

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 2, 2002, 22:45:52 ; Search time 2977.34 Seconds
(without alignments)
99.736 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaaaagtattccgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_sy:*
11: gb_un:*
12: gb_vl:*
13: gb_wi:*
14: gb_yi:*
15: em_fun:*
16: em_hum:*
17: em_in:*
18: em_or:*
19: em_ov:*
20: em_ph:*
21: em_pi:*
22: em_pl:*
23: em_ro:*
24: em_sts:*
25: em_sy:*
26: em_vl:*
27: em_wi:*
28: em_yi:*
29: em_htgo_hum:*
30: em_htgo_inv:*
31: em_htgo_rod:*
32: em_htg_hum:*
33: em_htg_inv:*
34: em_htg_rod:*
35: em_htg_other:*
36: em_htg_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	1604	1	MBU35021
2	18	100.0	453	1	MTU43540
3	18	100.0	9281	1	MHR3S1
C 4	18	100.0	9764	1	AF006921
C 5	18	100.0	10019	1	AF041819
C 6	18	100.0	12412	6	186264
C 7	18	100.0	17783	1	AE007028
8	16.4	91.1	1535	6	AK096713
9	16.4	91.1	1535	6	AK147694
10	16.4	91.1	16384	3	AF007160
11	16.4	91.1	42096	3	CEP21A3
C 12	16.4	91.1	42437	1	MTY436
C 13	16.4	91.1	48380	1	MTY13E12
C 14	16.4	91.1	4401	1	MTY13E12
15	16	88.9	4235	9	AK027649
16	16	88.9	4288	9	AK051548
17	16	88.9	4288	9	AK050515
18	16	88.9	184584	2	AL360171
19	15.4	85.6	1689	1	KPN252298
20	15.4	85.6	2154	9	B63424
21	15.4	85.6	2155	10	KNISK3A
22	15.4	85.6	2169	9	HIMG1YSYN
C 23	15.4	85.6	3357	10	HC004784
24	15.4	85.6	26857	9	AK048728
25	15.4	85.6	94707	2	AC010688
C 26	15.4	85.6	111084	9	AC006486
C 27	15.4	85.6	113688	9	AF005301
C 28	15.4	85.6	12169	9	AF008771
C 29	15.4	85.6	128077	9	AF000548
C 30	15.4	85.6	140072	8	H0512801
C 31	15.4	85.6	14236	2	AC069054
C 32	15.4	85.6	158861	2	AC025739
C 33	15.4	85.6	160270	33	AC025087
C 34	15.4	85.6	162507	9	AC090944
35	15.4	85.6	167108	2	AC068283
C 36	15.4	85.6	167760	2	CNS01DTN
C 37	15.4	85.6	186271	9	AC024168
C 38	15.4	85.6	186583	2	AC073094
C 39	15.4	85.6	189015	2	AC090560
C 40	15.4	85.6	190317	2	AC084332
C 41	15.4	85.6	223034	2	AC046147
C 42	15.4	85.6	237231	2	AC008482
C 43	15.4	85.6	250091	2	AL552185
44	15.4	85.6	280087	3	AE004526
45	15	84.4	4813	8	C00119864

ALIGNMENTS

RESULT	1
MBU35021	
LOCUS	MYcobacterium bovis BCG DNA flanking deletion region 3.
DEFINITION	27-APR-1996
ACCESSION	U35021
VERSION	U35021.1 GI:1049243
KEYWORDS	Mycobacterium bovis BCG.
SOURCE	Mycobacterium bovis
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium tuberculosis complex.
REFERENCE	1. (bases 1 to 1604)
AUTHORS	Mahairas,G.C., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE	Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis
JOURNAL	J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE	96200095
REFERENCE	2. (bases 1 to 1604)
AUTHORS	Mahairas,G.C., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.

TITLE Direct Submission
 JOURNAL Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology, Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
 FEATURES
 source Location/Qualifiers
 1..1604
 /organism="Mycobacterium bovis"
 /sub_species="BCG"
 /db_xref="taxon:1765"
 misc_feature 1400
 /note="site of deletion breakpoint in this organism. Virulent M. bovis has additional 10 kbp at this site"
 BASE COUNT 337 a 637 c 416 g 214 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 1; Length 1604;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 1 acatcaaaagtattcgcg 18
 ||||||||||||||||
 DB 465 ACATCAAAAGTATTCGCG 482
 RESULT 2
 MTU43540 3453 bp DNA BCT 14-AUG-1997
 LOCUS
 DEFINITION Mycobacterium tuberculosis rfbA, rhamnose biosynthesis protein (rfbA), and rmlC genes, complete cds.
 ACCESSION U43540
 VERSION U43540.1 GI:2326948
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (sites)
 Lee, T.
 Rhamnose biosynthetic genes related to a novel repeated sequence of Mycobacterium tuberculosis
 Unpublished
 Lee, T.
 Direct Submission
 Submitted (18-DEC-1995) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
 3 (bases 1 to 3453)
 Lee, T.
 Direct Submission
 Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
 Sequence update by submitter
 On Aug 14, 1997 this sequence version replaced gi:1304703.
 Location/Qualifiers
 1..3453
 /organism="Mycobacterium tuberculosis"
 /db_xref="taxon:1773"
 /clone="pIS116"
 150..1139
 /gene="rfbA"
 150..1139
 /gene="rmlB"
 150..1139
 /gene="rmlB"
 /note="previously known as rfbB"
 /codon_start=1
 /transl_table=1
 /protein_id="AAB66648.1"
 /db_xref="GI:2326950"
 /translation="MRLIVTGGAGFIGTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL

AIVEDAI RLVOGDITDAELYSOLVAESDAVVHFAAESHVDAVDNALDNPPEFLHTNVIGTF
 TILAEVRRHRLHHISTDEYVGDLELDRAFTETSTPNPSSPYSATKACADMVLVRA
 WYRSCVRATISNCNNYGPYQHVKEFIPROITNCLTGLKLYGGGNVYRWIHVDDH
 AIDPSRIYDELRCRIGETYLISSEGERDNLTVLSRLRLMDRDPDDFDHVTDRVAHDRLY
 AIDPSRIYDELRCWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
 150..1139
 /gene="rmlB"
 /codon_start=1
 /transl_table=1
 /product="rhamnose biosynthesis protein"
 /protein_id="AAB66647.1"
 /db_xref="GI:2326949"
 /translation="MKLIIVTGGAGFIGTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
 AIVEDAI RLVOGDITDAELYSOLVAESDAVVHFAAESHVDAVDNALDNPPEFLHTNVIGTF
 TILAEVRRHRLHHISTDEYVGDLELDRAFTETSTPNPSSPYSATKACADMVLVRA
 WYRSCVRATISNCNNYGPYQHVKEFIPROITNCLTGLKLYGGGNVYRWIHVDDH
 NSAVRILDRGRIGETYLISSEGERDNLTVLSRLRLMDRDPDDFDHVTDRVAHDRLY
 AIDPSRIYDELRCWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
 1136..1747
 /gene="rmlC"
 1136..1747
 /gene="rmlC"
 /note="previously known as rfbC"
 /codon_start=1
 /transl_table=1
 /protein_id="AAB66649.1"
 /db_xref="GI:2326951"
 /translation="MREPRELDVPGAEITPTIHVDSRGLFEMLTDHCEKAFAGHS
 LDVRQVNCYSAGYVRLGLHFAQLPQSOAKYVTCVSGSVFDDVVVDIREGSPYFGRWDS
 VIJLQDDRNIVSDGLAHGLIALQDQNSTWYLCASRYNPQRQHTICATDPTPLASIKRW
 SNGGQDPSISDRDAAPSFEEVRRGLLPQGTDTVSRLFEGMRGT"
 2202..2655
 /note="3-4 copies in M. tuberculosis genome; Found in the
 rfb cluster (inside the rfbA gene), also found in the mce
 of M. tuberculosis; present at 3-4 sites of the genome of
 H37Kv, H37Ra, Erdman, and many isolates; not found in M.
 smegmatis & M. avium."
 /rpt_type="dispersed"
 /rpt_unit="1806..2250
 BASE COUNT 679 a 1226 c 990 g 558 t
 ORIGIN
 Query Match 100.0%; Score 18; DH 1; Length 3453;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 1 acatcaaaagtattcgcg 18
 ||||||||||||||||
 DB 2216 ACATCAAAAGTATTCGCG 2244
 RESULT 3
 MBDR3S1
 LOCUS MBDR3S1 9281 bp DNA BCT 27-APR-1996
 DEFINITION Mycobacterium bovis deletion region 3, 5' end.
 ACCESSION U35017
 VERSION U35017.1 GI:1049238
 KEYWORDS
 SEGMENT 1 of 2
 SOURCE
 ORGANISM
 Mycobacterium bovis.
 Mycobacterium bovis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 9281)
 Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
 Molecular analysis of genetic differences between Mycobacterium
 bovis BCG and virulent M. bovis
 J. Bacteriol. 178 (5), 1274-1282 (1996)
 96200095
 JOURNAL
 MEDLINE
 REFERENCE 2 (bases 1 to 9281)
 Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.

```

TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
           Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES   Location/Qualifiers
           source          1..9281
                        /organism="Mycobacterium bovis"
                        /db_xref="taxon:1765"
misc_difference replace(1400..>9281,"")
BASE COUNT 1706 a 3191 c 2929 g 1453 t      2 others
ORIGIN

Query Match      100.0%; Score 18; DB 1; Length 9281;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgttcgqg 18
    |||||
Db 465 ACATCAAAAGTATTCGCG 482

RESULT 4
AE006921/c      9764 bp      DNA      BCT      27-APR-2001
LOCUS      Mycobacterium tuberculosis CDC1551, section 7 of 280 of the
           complete genome.
ACCESSION  AE006921 AE000516
VERSION     AE006921.1 GI:13879142
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 9764)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 9764)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Direct Submission
TITLE      Submitted (25-APR-2001) The Institute for Genomic Research, 9712
JOURNAL    Medical Center Dr, Rockville, MD 20850, USA
FEATURES   Location/Qualifiers
           source          1..9764
                        /organism="Mycobacterium tuberculosis CDC1551"
                        /strain="CDC1551"
                        /db_xref="taxon:8331"
                        /note="clinical strain"
                        122..1414
                        /gene="MT0095"
                        /note="This region contains an authentic point mutation,
                        causing a premature stop, and is not the result of a
                        sequencing artifact; similar to SP:PI6431 GB:X17506
                        PID:41684 PID:882614 GB:U00096; identified by sequence
                        similarity; putative"
                        1449..2123
                        /gene="MT0096"
                        1449..2123
                        /gene="MT0096"
                        /note="identified by glimmer2; putative"
                        /codon_start=1
                        /transl_table=11

gene
gene
CDS

gene
CDS

```

```

/product="hypothetical protein"
/protein_id="AAK44319.1"
/db_xref="GI:13879143"
/translation="MSYKHIAPSRVRLRQTRSTVYVKGSGSLSHRRVPTGIDGLAVNG
GREYRAVKPTGPGIQPKGDMMTVTVDAGPGRSVRSVEVAAPAAELPAIVADPRRIIR
ELIKSGTVRGNTKVPKALVCGSKFSTKMKLGLDPLRTISVTALRPNELVPSHPIGH
KWKWFFSLSTPTITRVETETDYHAAGAIKNGLKFPYFMTGFAKSNAAAGTEATLAKLSIDQ
YAKHKA"
2280..2873
/gene="MT0098"
2280..2873
/gene="MT0098"
/note="similar to GP:145427; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAK44320.1"
/db_xref="GI:13879144"
/translation="MDQPMNANIHVDALLDAMVPLGTQCVLDVCGDGLJAARLARRI
PYTAVDIDAPVLRRAQTRFANAPIRWLHADITMAELPNAGFDVAVSNAALIIHIEDTR
TALSRKGLGVTHGGTLAVVTWTPSLRNGLWHLTSWACGMANKVKGRWEISAPTKWP
PQYTLBELRSHVRALLPGACIRRLYGRVLYTWAPV"
3002..3772
/gene="MT0099"
3002..3772
/gene="MT0099"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNOKLRNRWELTICGLGGHVTYAPDAAALAKIKASTGIGE
VNRKLRGDFALAGGPGAGPAPDAPLIMRGKALROALIRALGVRLIRKALVIALAAW
AWRFKGAACALQATLIDRLPYLRAGCEKVDOMTVIHALFKALAAKPSHIALTCMLA
ATVILQKRGVGLWLLKWKGEYFAVVATSFIFLEVHILAKGITTTRVVTFSINVAAV
VTLISKRULFGVGRKKAYDVERGRQLDLERAAMIJ"
3791..4162
/gene="MT0099.1"
3791..4162
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MHGLKVVSKVTSAPFALETPARKAKAKARNIDJVAOCIKRTIDEG
VLAQCICGCSAKALPTAAWSROSQRPVATHCRPCCACGAAASALTCALALAWSVEE
VKAAPRAATESLQSRRAARESK"
4206..4973
/gene="MT0100"
4206..4973
/gene="MT0100"
/note="similar to GH:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine
nucleosidase/S-adenosylhomocysteine nucleosidase,
putative"
/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTYGVICAIPELAYLRGVLVDKAKROOVAIILFDSGLIDHR
VVIAAAGMGKVNVTGTTATLJADRFGRCTIVFTGVAGLIDPELCIGDIVIADRVVGHIF
GLITDERLRPYQPGCHIPFIPTEPRGCPYDVPDPAVIDRVKRLDQFTTLAPLSTAAGGGR
QPRITYYGTILTGIDYILHCERTNRRLHHELGGMVEMEGGAVAGICASFIDIPWILVIAL.
SDLAGADSGVDENRFVGEVNAASSARVLLRLLPVLTAC"
5105..7390
/gene="MT0101"
5105..7390
/gene="MT0101"

```



```

LLGDWRTRITELAAAGTAGLDTARALPAVTVRVCGAIGVIECDRPPVDLAVATPAA
LDRGVWLRPFRLNLYVAMPPICTPAETQITSAWVEVARLVGSLP"
3120..4280
/gene="biof"
CDS
3120..4280
/gene="biof"
/EC_number="2.3.1.47"
/note="Biof; similar to MTCY336.34c"
/codon_start=1
/transl_table=11
/product="putative 8-amino-7-oxononanoate synthase"
/protein_id="AAB96957.1"
/db_xref="GI:2791846"
/transl_table="MKAATQARIDDSPLAWDQVQRHEAGRLKRLPRPRAVATELD
LASNDYGLSRHPAIVDGGVQALRIWGAARSHASLVDPACHLSRAKRVVTPRHVDAAV
GLPSSCYTNLNAVGLSGPGSLIVSDARSHASLVDPACHLSRAKRVVTPRHVDAAV
AALKSREORAVVYVSDVSFNSDGLAPVRELIFVCRKHGALLIVDEANGILVGRGGNG
LXYELGAGAPDVVMTTTLKALGSGGVGLGPPVPRKHLIDRAKPIIFDTGLAPAAV
GAARALRLQAEPQAAVQVNLNHAGELARMCVAAVPSAMSVIIGEPESAVAAAAA
CLDAGVKGCFRPPTVPAGTSRLRTASINAGELELARRVLTDLAVARR"
4277..4957
/gene="biob"
4277..4957
/gene="biob"
/EC_number="6.3.3.3"
/note="Biob; similar to MTCY336.33c"
/codon_start=1
/transl_table=11
/product="putative dethiobiotin synthetase"
/protein_id="AAB96958.1"
/db_xref="GI:2791847"
/transl_table="MTILVVTGTGTGKTVVCAALASAAKAGQIDVAVCKPVGTGTA
RGDDDLAEVGLAGVQTQGLAGRYPOPMAPAAAAHACMALPARIQIVRLITADLRPG
RLTLVEGAGLLVELAPGVTLRDVAVDVAVALVVTADLTGLTNHFKITLFEALAAQ
VSCAGLVGSGWPDPPGLVAASNRSALARLATVRAALPAGAASLDAGDFAAMSAAAFDK
NWAAGLVG"
4957..5466
/note="ORF6; similar to MTCY336.32c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96959.1"
/db_xref="GI:2791849"
/transl_table="MWISIELVFUSYTEAAIRRIWAGLILAAALHPSQAVASRPHYSIAV
AERTAPEDEPLGAVARKIPLDGVIGAPVIGRANVVTPLVVPSTELLALHAKVHRL
CGPHLAPAPMANSLPGOWTAHVTLARVGGHOLGKALKIACGRPSKILRRFAGLRWNG
NTRAEYLLIG"
complement(5612..6478)
/note="ORF7; similar to MTCY251.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96960.1"
/db_xref="GI:2791849"
/transl_table="MDWLHPDGLTUTERAKRGTITLSNOOYDGMSKI.SGYLTPOARA
TFEAVLAKAAPGATNPDDHTPVITTPDAAAIDRDTKSAQRNIHUGILAGLRALIAS
GELGONGPLSVIVVTTLTDLQTCAGKGTGCGTLLPMADVIRMTSHAHYSPASGR
YPOAIFDGTFLALHTRKLASPAQRIMLFANDRGCTKPGCUAPASQAHYVGTWIS
TGRTDITDLTLACDPDNRKLAEGWTRKNTHGHTENLPPPHLDHCQPTNFFHHEKL
LRHNEDNHDOP"
complement(6339..7007)
/note="ORF8; similar to MTCY336.16"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96961.1"
/db_xref="GI:2791850"
/transl_table="MLANSREELVEFDALDAELDRLDEVSFFVLTTPFLRLSLERLE
CLVRLPVGHTLNLQDLOQASEEELGCTLCCALANRLITKPDAAALKJADAADLGP
RALTCPEPLAQOLATATAQROGLICEAHIKVIIRALFRPARRGCGVHPHGRSRPGRQ
SRSSSRAGRLRPAGHGLATPKRRPHRHTKPTQRHHHPQPAIRKHVTAKWLDPPPS
AghL"
7455..8504
/gene
```

```

/gene="biob"
7455..8504
/gene="biob"
/note="Biob; similar to MTCY336.15c"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAB96962.1"
/db_xref="GI:2791851"
/transl_table="MTQAAVUKPTNDAGDGGNNSDILVAVROCVLQKFGCINOTQVLA
VIALPHUHLRELLALAHFVHKWCCGFEVEVGGILSLKTCGCCPHICHPCUSGCFASPV
KSAWLDIPSIVEAAKQATANKATEREILVAVARGPDEKLMQVAAAGIEAIRKEVEINIA
CSIGMLTAEQVDQDLAAKGVHKNINIETARSEFFANVVVTHITWEERWOTLSMVRDAGME
VCCGTLIGMETIQQRAEFAAEALAEIGPDEVPNLNPREGTTFADI.EVMPVGDALKA
VAAPRLALPKTLMKRFAGRGKLTCLDCLCAKRGILGGINAVIVGNVLTITICHPAEADLEL
LHLOMPILKANASHL"
8825..9406
/note="ORF10; similar to MTCY336.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96963.1"
/db_xref="GI:2791852"
/transl_table="MUGLSATGVLCGLWAIAPPJIAHVVAITRAGERVIEYLGSESQ
NFIAPMLIGLILSVIAYVASALMKWQRHRRGQPMVAGLSIGLTTAAAIAGVGALVV
RLKYGALDFDTPLSKGDHIALTYTQAPPVFFRRGPIQIALTLMPAGTASLVYALLA
ACTHARDLGGYPAVDPSNNAKTEALETTPQAPVS"
complement(9571..>10019)
/note="ORF11; similar to MTCY336.12"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAB96964.1"
/db_xref="GI:2791853"
/transl_table="RMAGROMGDFLHPELEDILSTPEVSHVFGDTKICSAVPTPPVLI
VQAVHDTLIQVSDIDALADSYTAGGANVTYHRDLFSHVSILHPI.SAPMTLKWLTDRFA
GKPLTOHVRVETMTLTPMTYAGMARLAVIAAKVITGRKLSRRPL"
BASE COUNT 1476 a 4134 c 4655 g 1774 t
ORIGIN

Query Match 100.0%; Score 18; dB 1; Length 10019;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conserved Ives 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatcaagatgattcagcag 18
|||||
db 6619 ACATCAAGATGATTGCG 6602

RESULT 6
LOCUS 186264 12412 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 18 from patent US 5700683.
ACCESSION 186264
VERSION 186264.1 GI:3205982
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12412)
AUTHORS Stover,C.Kendall and Mahairas,G.G.
TITLE Virulence-attenuating genetic deletions deleted from mycobacterium
BCC
JOURNAL Patent: US 5700683-A 18 23-DEC-1997;
FEATURES location/Qualifiers
source 1..12412
/organism="unknown"
BASE COUNT 2146 a 4274 c 3875 g 1917 t 200 others
ORIGIN

Query Match 100.0%; Score 18; dB 6; Length 12412;
```

Best Local Similarity 100.08; Pred. No. 14;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 acatcaagtgattcg 18
|||||

Db 468 ACATCAAGTGATCGC 485

RESULT 7

AE007028/c

LOCUS

DEFINITION

Mycobacterium tuberculosis CDC1551, section 114 of 280 of the complete genome

AE007028 AE000516

AE007028.1 GI:13881250

Mycobacterium tuberculosis CDC1551.

Mycobacterium tuberculosis CDC1551

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Mycobacteriaceae;

Mycobacterium; Mycobacterium tuberculosis complex.

1 (bases 1 to 17783)

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

Unpublished

2 (bases 1 to 17783)

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

Direct Submission

Submitted (25-APR-2001) The Institute for Genomic Research, 9712

Medical Center Dr. Rockville, MD 20850, USA

Location/Qualifiers

1. 17783

/organism="Mycobacterium tuberculosis CDC1551"

/strain="CDC1551"

/db_xref="taxon:83331"

/note="clinical strain"

complement(104..826)

/gene="MT1622.1"

complement(104..826)

/gene="MT1622.1"

/note="identified by Glimmer2; putative"

/codon_start=1

/product="conserved hypothetical protein"

/protein_id="AAK45890.1"

/db_xref="GI:13881251"

/translation="MLAKLAPGATNPDDHTPVIDTTPDAALIDKDKSQAKRNDGL
LAGRLATASGLGQHNGLPVSIVVTTLTDLTGAGKGTGGTLLPMADVIRMTSH
AHYSPASGRVPAIFDHTGTPLAYHTRLASPAQRIMLFANDRCCTKPCDAPYHS
QAHVHTASTGRTDITDLTACDPDNRLAEGWTHKTHGHTWELPPPHLDHGQPR
TNTFFHKKLRHNDNDNDHP"

complement(831..1499)

/gene="MT1623"

complement(831..1499)

/gene="MT1623"

/note="identified by Glimmer2; putative"

/codon_start=1

/product="conserved hypothetical protein"

/protein_id="AAK45891.1"

/db_xref="GI:13881252"

/translation="MLANSRELVEFDALDAELDRLEVSFEVITTHPERLSRLERLE
CLVRRLPAVGHTLNOLDTQASEEIGTCCALANKRIITKPAALAKIAADALGPR
RALTGEPALPOLTATATAQROGLTGFAHIKIRALRIPPARRKGCYVIIIPIGRKSRPGRO

gene

CDS

1947..2996

/gene="MT1624"

1947..2996

/gene="MT1624"

/note="similar to GP:1680650; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=11

/product="biotin synthase"

/protein_id="AAK45892.1"

/db_xref="GI:13881253"

/translation="MTQAATRTNIAGODGNNSDILVYAKQVQLQKQEGILNQDOVLA
VIQLPDDRLREELALAHIEVRMRWCPEVEGIIISLTKGCPEDCHFCQSGLPASPV
KSAWLDITSLVENAKQKTSATGATEFCIVAAVRGPDERLMAQVAGLIEAIRNEVEINIA
CSIGMIATAEQVLDAAAGTAVIRYNNHLETSRFFANVTTHWEERWOTISMYVDAGME
VCCGCIICMGPTLQAKAEFAELAFGLPDEVLNPLNRPGTTFADILVMPVGDALKAL
VAARIALPRLPTMLKPFAGGREITLGLDGAQRGLGGINAVIVGNYIITFLGRVAEADLEL
LDELQMLKALNASI"

2997..3246

/gene="MT1625"

2997..3246

/gene="MT1625"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45893.1"

/db_xref="GI:13881254"

/translation="MVEIVAGKQRAIVAAGVYVNYVTGELATATPTAAKMLLEPRFC
AKGRMVVGVKPIQGWAKCSHHQGVDSADIATQR"

3317..3898

/gene="MT1626"

3317..3898

/gene="MT1626"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45894.1"

/db_xref="GI:13881255"

/translation="MIATISATGVLVGLWMIAPPVIAVVAITRACERVHHVYKGSQ
NFEIAPMLAGLISVLAVVSSALMWREHRCPOWVAGLISGLTTAAALAACVGCALVV
ELVGCALPDLTIVLSKSHALTYVIGAPVFAPRPLQIALIMVAVGLASLVYVALLA
AALAHDDAGTAAVDFSSNAKTEALHTTQAPVS"

complement(3953..4054)

/gene="MT1627"

complement(3953..4054)

/gene="MT1627"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK45895.1"

/db_xref="GI:13881256"

/translation="MRSATGTANSDLAHLNLYTDLNQFDLPLESKR"

complement(4063..5430)

/gene="MT1628"

complement(4063..5430)

/gene="MT1628"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45896.1"

/db_xref="GI:13881257"

/translation="MPDHEGSLMVEPGLAGATGAETGRPPHEFLQKVRPLPSPD
DPFYPPAGYQAHVPGTVLRSRDVELAFMGLIPQVATQLLYRTTNMGNPEATVTT
VIVPAELAPGQTCPLLSYQCADIDANSSRCFYSYALRRKALGSLTOWELIMISAALA
EGWAVSYDPGPKLWSPYEPGYRVLDDGIRALNSRVLSPATPGLIWAALQSHYIP
ESAWAAACGYPADILDTVGAVGSPVLDGLTFRRLNGTLIAGLPAIIVAAIQHSYIP
GLIARYIKHANDEGROIIEQLTEMTTVDVIRMGARDMGDFIIDEPLDIIISTPYVSHV

gene

CDS

1947..2996

/gene="MT1624"

1947..2996

/gene="MT1624"

/note="similar to GP:1680650; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=11

/product="biotin synthase"

/protein_id="AAK45892.1"

/db_xref="GI:13881253"

/translation="MTQAATRTNIAGODGNNSDILVYAKQVQLQKQEGILNQDOVLA
VIQLPDDRLREELALAHIEVRMRWCPEVEGIIISLTKGCPEDCHFCQSGLPASPV
KSAWLDITSLVENAKQKTSATGATEFCIVAAVRGPDERLMAQVAGLIEAIRNEVEINIA
CSIGMIATAEQVLDAAAGTAVIRYNNHLETSRFFANVTTHWEERWOTISMYVDAGME
VCCGCIICMGPTLQAKAEFAELAFGLPDEVLNPLNRPGTTFADILVMPVGDALKAL
VAARIALPRLPTMLKPFAGGREITLGLDGAQRGLGGINAVIVGNYIITFLGRVAEADLEL
LDELQMLKALNASI"

2997..3246

/gene="MT1625"

2997..3246

/gene="MT1625"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45893.1"

/db_xref="GI:13881254"

/translation="MVEIVAGKQRAIVAAGVYVNYVTGELATATPTAAKMLLEPRFC
AKGRMVVGVKPIQGWAKCSHHQGVDSADIATQR"

3317..3898

/gene="MT1626"

3317..3898

/gene="MT1626"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45894.1"

/db_xref="GI:13881255"

/translation="MIATISATGVLVGLWMIAPPVIAVVAITRACERVHHVYKGSQ
NFEIAPMLAGLISVLAVVSSALMWREHRCPOWVAGLISGLTTAAALAACVGCALVV
ELVGCALPDLTIVLSKSHALTYVIGAPVFAPRPLQIALIMVAVGLASLVYVALLA
AALAHDDAGTAAVDFSSNAKTEALHTTQAPVS"

complement(3953..4054)

/gene="MT1627"

complement(3953..4054)

/gene="MT1627"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK45895.1"

/db_xref="GI:13881256"

/translation="MRSATGTANSDLAHLNLYTDLNQFDLPLESKR"

complement(4063..5430)

/gene="MT1628"

complement(4063..5430)

/gene="MT1628"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45896.1"

/db_xref="GI:13881257"

/translation="MPDHEGSLMVEPGLAGATGAETGRPPHEFLQKVRPLPSPD
DPFYPPAGYQAHVPGTVLRSRDVELAFMGLIPQVATQLLYRTTNMGNPEATVTT
VIVPAELAPGQTCPLLSYQCADIDANSSRCFYSYALRRKALGSLTOWELIMISAALA
EGWAVSYDPGPKLWSPYEPGYRVLDDGIRALNSRVLSPATPGLIWAALQSHYIP
ESAWAAACGYPADILDTVGAVGSPVLDGLTFRRLNGTLIAGLPAIIVAAIQHSYIP
GLIARYIKHANDEGROIIEQLTEMTTVDVIRMGARDMGDFIIDEPLDIIISTPYVSHV

FGDTKLGSVPTPPVLIQVAOAVHDYLDVSDIDLADSVYTAGGANVTYHRDLFSEHVS
HPLSAPMTLRWLTDREAGKPLTDHVRVTWPTIFNPMYAGMARLAVIAAKVITGRKL
SRRL"

gene complement(5660. .6193)
/gene="MT1629"
CDS complement(5660. .6193)
/gene="MT1629"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45897.1"
/db_xref="GI:13881258"
/translation="MTSSVRRQLAEKVDLRELAHQEAVFSDPHRLPGIRMIASVYL
GVVSPATPELPADTRWHPVSSLPMAFDGPMVTHARTLIAKMSVNTICFALAPKE
FALSTLDIYGAALGYQVDATNQLQVLRARRVITQTGTIAQSGSGCKHPAALRYFTDS
QLRVTFDEFAALRPQGL"

gene 6419. .7468
CDS /gene="MT1630"
6419. .7468
/gene="MT1630"
/note="similar to SP:P11458 GB:X12713 PID:581137 GB:U00096
PID:1651334; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="quinolinate synthetase"
/protein_id="AAK45898.1"
/db_xref="GI:13881259"
/translation="MTVLNRTDTLVDELTAADITNPLGCGVGDDGDEWAAETRLRLAHL
RGATVLAHNYQLPAIDQADHVDGDSLARSVAAPEDTIVFCGVHFMETAKILSPH
KTVLPDQAGCSTADSIPTDELRAWKDEHFGAVVSYVNTAAVKALTDICCTSSNA
VDVASIDPDEVLFCPDQFGLAHVRYTRKNIHWAGCEIVHAGINGDELAQARA
HPDARFVHPGECATSAIYLAGEGAFPAERVKILSTGCMLEFAAHTTRAKOVLVATEV
GMLHOLRAAPEVDPRAYNDRASCYKMKITPAALLKCLIVEGADEHVHVDPGIAASGR
SVORMIECHPGGGE"

gene 7468. .9051
CDS /gene="MT1631"
7468. .9051
/gene="MT1631"
/note="similar to GP:608530; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AAK45899.1"
/db_xref="GI:13881260"
/translation="MAGPAWRDAADVVVIGTGVAGLAALAADRAGRSVVVLSKAAQT
HVTATHYAGGGLAVVLDPNDSVDHVDTLAAGAGAGLCPDAVYSIVADGYRATVDLV
CAGARLDESVPGRWALTREGHSRRRIHVAGGDATCAEVORALODAPAGMLDIRTCHVA
LRVLHDGTAFTGLLVVRPDGCGIISAPSVILATGGLGHLIYATNPAGSTGDGTALGL
WAGVAVSDEFTIQFHTMLFAGRAGRRPLITEAIRGEGAILVDROGNSITAGVHPMG
DLAPRDVVAALIDARKLKATGDPVYLDARGIEGFASRPTVTASCAAGIDPVRQ?IP
VYPGAYSCGGITVDYGTOTELLGLYAAAGEVARTLHGANKRLASNLIEGLVYVGRAG
KAAAHAAAGRSRATSWPEPISYLTALDRGLQRAMSRDASHYRAAAGLHRLCDS
LSGAQVRDVACRRDFEDVALTLVAOSVTAALARTESRGCHHRAEYPCTVPEQARSIV
VRGADANAVQVALVVC"

gene 9021. .9908
CDS /gene="MT1632"
9021. .9908
/gene="MT1632"
/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262
PID:456041; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11

Query Match 100.0%; Score 18; DB 1; Length 17783;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18
|||||

DB 1111 ACATCAAGTGATTCGCG 1094

RESULT 8
AR096713 LOCUS AR096713 1535 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6008201.
ACCESSION AR096713
VERSION AR096713.1 GI:10025749
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1535)
TITLE Riley,L.W.
FEATURES DNA molecule encoding for cellular uptake of mycobacterium
tuberculosis and uses thereof
JOURNAL Patent: US 6008201-A 1 28-DEC-1999;
FEATURES location/Qualifiers
source 1..1535
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 6; Length 1535;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18
|||||

DB 841 ACGTCAAGTGATTCGCG 858

RESULT 9
AR147694 LOCUS AR147694 1535 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6224881.
ACCESSION AR147694
VERSION AR147694.1 GI:15111784
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1535)
TITLE Riley,L.W. and Chong,P.
FEATURES DNA molecule fragments encoding for cellular uptake of
Mycobacterium tuberculosis and uses thereof
JOURNAL Patent: US 6224881-A 1 01-MAY-2001;
FEATURES location/Qualifiers
source 1..1535
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 6; Length 1535;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18
|||||

DB 841 ACGTCAAGTGATTCGCG 858

RESULT 10
AE007160 LOCUS AE007160 16384 bp DNA BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 246 of the
complete genome.
ACCESSION AE007160 AE000516
VERSION AE007160.1 GI:13883389
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551.

Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS

1 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.

Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains

JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.

Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source

1. 16384
Location/Qualifiers
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"

gene

CDS

96. .839
/gene="MT3557"
/note="similar to SP:000298; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47897.1"
/db_xref="GI:13883390"

871. .1566
/gene="MT3559"
/note="similar to SP:000298; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47898.1"
/db_xref="GI:13883391"
/translation="MGAGALITAVLLIAGVMTIIVAFADGCPHAEVTFARGTEPPP
GIRVGQAFVDSLRQQTGMEIGVYPNYAASRLQLHCGDCANDAI SHIKSMASCPNT
KLVLGYSOGATVIDIVAGPLGSI SFGSPGLAAADNVAAVAFVFNISNRAGGSLS
LSPFLGSKAIDLCNPTDPICHVCPGNEFSGHIDDIPTVTYTTQAAASFVQRLRAGSVPH
LRGSPQLPGSVLQMPGTAAPAPESLRG"

gene

CDS

871. .1566
/gene="MT3559"
/note="similar to SP:000298; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47898.1"
/db_xref="GI:13883391"

1624. .1863
/gene="MT3560"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47899.1"
/db_xref="GI:13883392"
/translation="MRFICVIPPPOHSGRWAGARRLTSILVAAFAAATLLTPAL
APPASAGCPDAEVWFARGTGERPGLRGVQAFVSSLROOTNKSICTGYGVNYPANGDEL
AADGANDSDHIQQNASACRATRLVGGT SQGAVIDIVTAAPLPGIGFTQPLPAA
DHIAAIALFGNPSGRAGGLSALTPQFGSKTINLCNGDPICSDGNRRRAH/GYVPG
MTNQAAAFVASRI"

gene

CDS

1624. .1863
/gene="MT3560"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47899.1"
/db_xref="GI:13883392"
/translation="MTAADWHYAHCLMLAPDAPQARRPSTQPNAHLSAECAAMNRN
HFTGREVCGHANKTLGLTSTAALRRVDAGRHNKQ"

gene

CDS

1835. .3523
/gene="MT3561"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47900.1"
/db_xref="GI:13883393"
/translation="MMPGVJTNSESPTAADHDIRTATRETLEDYTLRLAPKSYRRWPP
AVGIALSGGIAYIADFAIGANVGITWGTANALGCIATFALYVVPVGLPIAYAAAYN
IIDLITRSGSGGYGYSVTVNIFATFTTFFALEGSIIMAGCIKILILPIIWAAGYACS
TILITPLVYVGMKVISQIOWMTPLWLIMAAPFGYVVSHPDSTGQFFSYAGDGHG
GISGSVLLAAGVILSLIAQIAEQIDYLRFPPTPENAKWMTWTILAGHWAFGA
TKQITIGFLAVILMANIPGSSITIANQPIYFVPCWMLATLAVILVLSQIK
INVTNAYSCSIANTNSFTRI TKHYIGRVVFLGVNLAIALLMEANMPDF,NTIILCCYA
NCGMAVVAASDIDGNKYLGLSPKTPFERGMLYAINPVCGSILAAAGLSIVTFE
CGICAAQOPSPYALVVALVWPPILAAATKGKYLRTTHDCIDLPMDYDEHGNPSAAV
LTCVCHODFERPDMACUTCHAHVCSLCLSTDKQAEHVLPGLAKAHIPGDOVP"

CDS

1835. .3523
/gene="MT3561"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47900.1"
/db_xref="GI:13883393"
/translation="MMPGVJTNSESPTAADHDIRTATRETLEDYTLRLAPKSYRRWPP
AVGIALSGGIAYIADFAIGANVGITWGTANALGCIATFALYVVPVGLPIAYAAAYN
IIDLITRSGSGGYGYSVTVNIFATFTTFFALEGSIIMAGCIKILILPIIWAAGYACS
TILITPLVYVGMKVISQIOWMTPLWLIMAAPFGYVVSHPDSTGQFFSYAGDGHG
GISGSVLLAAGVILSLIAQIAEQIDYLRFPPTPENAKWMTWTILAGHWAFGA
TKQITIGFLAVILMANIPGSSITIANQPIYFVPCWMLATLAVILVLSQIK
INVTNAYSCSIANTNSFTRI TKHYIGRVVFLGVNLAIALLMEANMPDF,NTIILCCYA
NCGMAVVAASDIDGNKYLGLSPKTPFERGMLYAINPVCGSILAAAGLSIVTFE
CGICAAQOPSPYALVVALVWPPILAAATKGKYLRTTHDCIDLPMDYDEHGNPSAAV
LTCVCHODFERPDMACUTCHAHVCSLCLSTDKQAEHVLPGLAKAHIPGDOVP"

gene

CDS

3485. .4378
/gene="MT3562"
/complement(3485. .4378)
/note="MT3562"
/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U00096
PID:1788657; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tRNA pseudouridine synthase A"
/protein_id="AAK47901.1"
/db_xref="GI:13883394"

3485. .4378
/gene="MT3562"
/complement(3485. .4378)
/note="MT3562"
/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U00096
PID:1788657; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tRNA pseudouridine synthase A"
/protein_id="AAK47901.1"
/db_xref="GI:13883394"
/translation="MSLTKRPKSPORPRISGVVRLRLDIAYDGTUFAKWAQVCG
KTVAGDIDDAULTTTPRTVRLAAGRTDAGVHAGQVAHVDPADALPNAYPRAGHVG
DPEFTPLRLRLRGKFLPADVRLDITTRAPAGEDARFSALRRIYVRLSTAYCYGVEPQA
RYFTLMPRELIDIDAMTAASHDLMLGIIIDFAFCRHREGATTIRIQRLIMSRAGKTLVTA
IIVTADAFQWSMVRSILVGAIIAVGEHRRATTWCRELLTATGRSSDFAVAFHGLTILQV
IYPPIDQILASRLVTRDVKRG"

gene

CDS

4323. .4865
/gene="MT3563"
/complement(4323. .4865)
/note="MT3563"
/note="similar to GB:J01685 SP:P02416 GB:X00766 PID:147716
PID:42800; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L17"

4323. .4865
/gene="MT3563"
/complement(4323. .4865)
/note="MT3563"
/note="similar to GB:J01685 SP:P02416 GB:X00766 PID:147716
PID:42800; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L17"
/protein_id="AAK47902.1"
/db_xref="GI:14883395"
/translation="MLKPIKGPRIARSSSHQKAILANLATSILFHCRTTTPPKARAL
IKYAKLLITHAKKALINKREVLKKLRDQVVIITLFAEIGDPAIDKACGYTRILKIEA
IKGHNAPMAVLELVKEKIVTSEANKARVAAQAQAKKAAAMPTESEAKIYAEFGDDV
CASEPDAAKAPERPPTPEAPN"

gene

CDS

4897. .5940
/gene="MT3564"
/complement(4897. .5940)
/note="MT3564"
/note="similar to SP:P72404; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase, alpha subunit"

4897. .5940
/gene="MT3564"
/complement(4897. .5940)
/note="MT3564"
/note="similar to SP:P72404; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase, alpha subunit"
/protein_id="AAK47903.1"
/db_xref="GI:13883396"
/translation="MLISORPTLSEDLVLTNRSOFVIEPLPFGCTIGNSLRLTLLS
SIPGAATVSTRIDGVHEFTVTPGVKQEDVTEILLNLSLVVSSEIEHIVTMYLKKQGP
GHVTAGDIIPPPAGVTVHNPQHIAITLNDKRLKLELVVERGHGYVAVQNRASGAIEIG
RIPVDSIYSPVLKVTYKVDATREQRTDFKLILDVETKNSISPNDALASAKTIVLEI.
FGLARELNVAEGIEIGSPAEADHIAFALPIDDLDLTVSYNCKIKREGVHTVGLV
AYDEODYAETEOL"

gene

CDS

6092. .6697
/gene="MT3565"
/complement(6092. .6697)
/note="MT3565"
/note="similar to SP:P81288; identified by sequence
similarity; putative"

[illegible]


```

/gene="hycQ"
1629. 3095
/note="hycQ"
/note="Rv0086, (MTCY251.04), oxidoreductase, len: 488,
hycQ, most similar to sp|P77437|HYFE_ECOLI HYDROGENASE-4
COMPONENT F(EC 1.- (526 aa), FASTA scores, opt: 948
z-score: 1117.4 E(): 0; 35.9% identity in 493 aa overlap.
Also similar to E. coli d9087711 & NUOL_ECOLI P33607 nadh
dehydrogenase chain 1 (613 aa) FASTA scores, opt: 360,
z-score: 354.9, E(): 3.2e-13, (27.9% identity in 488 aa
overlap), and to NUON_ECOLI P33608, nadh dehydrogenase 1
chain n (425 aa), FASTA scores, opt: 375, z-score: 371.4,
E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/transl_table=11
/transl_table="MTGLLLAAILAPIAASIASLITQWRRTATLTATSATVLAACAV
AMGFMGSGAGLGLGLRADALIVMLVIGVGTLTATASTIGITDELTAHGLDGR
SARLVGLTAPFLCAMLAVLANCANNIGVIAEATVITFLVGRIRRTTALEATWTK
VIGSVGTAVFGLTGLVFAARDSGAAGALNLLDLAEHAAGLDPGARLAGLLY
IYCGAKAGLPFPHWTLADHSAQAPAPVMSGLVAVFSLVLRKRP ILDAVSGPAY
LRLNGVLGLATLLVAVLMLVTQDVKRLAYSSMEHMGILIAAAGACTTIAIALLL
HYLAHGIGKTLVFLAGGQQAHDSTADITGYMRRSLRGVSHFAGLVILGLPPF
AMFASLEIAKSLANERLWAGRAALLDIAIGFTALARNSSRMLLGTTPAACAPAITVP
ATAAALMVGIVVSAALGITAGPLADLLGIAASNVGLP"
3092. 4570
/gene="hycE"
3092. 4570
/gene="hycF"
/note="Rv0087, (MTCY251.05), oxidoreductase, len:
492, hycE, similar to HYCE_ECOLI P16431 formate
hydrogenlyase subunit 5 (569 aa) FASTA scores, opt: 680,
z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa
overlap) and to NUOD_ECOLI P33600 nadh dehydrogenase 1
chain d, (407 aa) FASTA scores, opt: 245, z-score: 293.1,
E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/transl_table=11
/transl_table="MMSASWLRIHVVSERGLIATAEQIWAQSPFLALVAAHDDGDSLRV
VILFAGYDPRRVEVYVADNPETRSIAYLSFPAGREFREMAIDYIIRPVGHPKPR
LVRHAWHDNPMPTDAGPAPFTHGCPFAVAGIIPGVVHAGIIEPCHF
RFSVAGETIVRLKARLWVPHRGIEKLEHCPAPAAVDIAERISGPTSAHAALHSIAI
EDALGETPHEVRLRLVLEKRIENHAADLCALANDGVYSLANAHOVR IENLILRR
NAATVGRILLRGLRAGGVALRALPIIDELAAADLAEVLAETLANSVYDRFAGTA
VLHPDDASALGCLGYVARSGLSDRDEVEITVLTPEDGPDGVLAIVYVRDEF
AASALAQHWESHTGPIEVAATLHPVGAPSSGIVGEMRGITVIRKVEIDVDGRTIR
AKVVDPSFNWPAIPVAMADTIIVDPKPLANKSFNQSGANDL"
4605. 5279
/gene="Rv0088"
4605. 5279
/gene="Rv0088"
/note="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/transl_table=11
/transl_table="MSVYKAPSRVRLQRTSTVWKGSGSLSHRRVRVIGDGLAVMG
/transl_table="MSVYKAPSRVRLQRTSTVWKGSGSLSHRRVRVIGDGLAVMG
GREGYRAVKPCTGPIQPGDMVTVVVDAGPGRVSRVSEVAAFAELFAIVADPRHR
ELDGSYVRNKKVPAKLVGSKFSTKMKLGLPYRTSRVTKALPNELVEWSHPLGH
RWRWEFSLSPTLTRVETEDYHAAGAINKGLAFYEMTGFAKSNAGIETLAKLSQD
YARGRA"
5436. 6029
/gene="Rv0089"
5436. 6029

```

(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBpase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source

Location/Qualifiers

1..43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"

source

<1..1481
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y77"

gene

78..1481
/gene="Rv3448"

misc_feature

78..164
/gene="Rv3448"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"

CDS

78..1481
/gene="Rv3448"
/note="Rv3448, (MTCY77.20), len: 467 aa. Unknown membrane protein, contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature. Some similarity to AL021930|MTV035_18 Mycobacterium tuberculosis (472 aa) fasta score, opt: 429 z-score: 488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv3448"

/protein_id="CAB08736.1"

/db_xref="GI:3261767"

/db_xref="SPTREMBL:O33354"

/translation="MPTSDPGLRRVTIAGAAQAVH.TLPAAVHPVATLIPSLVDIAGPR CASPATARYOLSGALPAPNATLAUCGIRDCGAVLVLIKSSAPPTPRCDQVAVAV AAADLTAPQCQRTFRLSGAASCLTAGGELMLVKNAGTNTVTRYSIATAGVAAA GLAALLFAVICTRTYRPIAGLTLSVATITFGAVAGILAVPGVGVHSLVAAQMAAA TSVLAMRTGGGTLTAVACCAVVAATILVGATITAPVPIUSIATLASFGLLEVS ARMAVLLAGLSPLRPALNPDDADALPTDRLTTRANKADAWLTSILIAAFASATIGA IGTAVATGCIHRSSMGGTALAAVTGALLLLHARSADTTRRSILVFAICGTTTATATTA ADRALEHGPWTAALTAMLAAMFGLGVAPALSLSPTVYRTIELLECIALTAMVPLTA WLCGAYSAVRHLDLTWT"

611..>43401

/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

/clone="Y13E12"

1478..2845
/gene="Rv3449"

1478..2845
/gene="Rv3449"

/note="Rv3449, (MTCY13E12.02), len: 455. Function: probable precursor of serine protease. Has putative signal peptide N-terminus and hydrophobic stretch at C-terminus. Contains three signatures typical of family: aspartic acid active site (PS00136), histidine active site (PS00137), serine active site (PS00138). FASTA results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E(): 8e-07; (38.0% identity in 387 aa overlap), similar to MTCY15F10.29 (45.7% identity in 451 aa overlap)"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv3449"

/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:Q06316"
/translation="MTTSKTRLLRVVLSALATLSGLGTPVAHANVSPPPIDRWLPESAL IAPRPPTVQREVCTEVTAEASGRAERSAQLADLDQVWRLTRGACQKRVAVIDTGA VHKRLPKVAVAGDYYFTGDTGADCAHGTIVAGIJAAPDAQSINFSCVAIDWTLI1S1 KQSSKFAFVGDIPSTGVDYDTMAKAVKTAADLCASVINISSIACVFAAAPPDRAI CAALAYADVKNVIVAAAGNTGGAACPPPOAGVTVTJDSVTVAVS/PAWYIDYVITVGS VNAQEPSAFTLACHVWLVYAAATGEAVTSLSPFGDCTVNRLLGQHSI1F1SCTSYAAPV VSGLAALIKARPPHITAKQVMORIESTAHIPPAWHPLYGVNCTVDAALAAVSSSLPOA CTAISDPAFAVAVLPVPRNSTPGFSDDKRLHTAFAGAAICLLALMATIAYASRKLPRGRN GIAGI"
1757..1789
/gene="Rv3449"
/note="PS00136 Serine proteases, subtilase family, aspartic acid active site"
1862..1894
/gene="Rv3449"
/note="PS00137 Serine proteases, subtilase family, histidine active site"
2456..2488
/gene="Rv3449"
/note="PS00138 Serine proteases, subtilase family, serine active site"
complement(2810..4222)
/gene="Rv3450c"
complement(2810..4222)
/gene="Rv3450c"
/note="Rv3450c, (MTCY13E12.03c), len: 470. Function: unknown but similar to Rv3895c MTCY15F10.17 (FASTA score: 1.6e-27; 36.4% identity in 475 aa overlap): possible membrane spanning region near N-terminus, and to Y14967|MLH628_16 Mycobacterium leprae cosmid H628; (481 aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 42.9% identity in 480 aa overlap, also similar to Rv3869 (MTV027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:Q06317"
/translation="MPSVATWIVNSGYRFLRRIRFALLPGLWCAATGALNARTTSL ALGVLAIVAAKRAALVALLRQSGALQAPIVMGRESGALYVYKVIDVWHLVNLASAR LIAATHANIPVPSSEELAHKRGPIILIGTAPQLLIQPLAGAESAMAIQDSNGSGSTT VVGFAFSSSAVLIAGEHILVATESGSPTYLYGGRRAVVIJAHVAVWALRQGRV DIIWVAUSLLNAVIEADPIAPRIKRGGRASVGLPGLVGGVYVRIKASGDEYVVLIED GYGRIGQVAADLLKPVUSGCSYNVITVAPDVTVPVIVNTLPVSAFPDPPHIVDGGSP GHAVTTLCTWTPAOCGAARVAFIAGSGPPVPLGCVPTLQAQIDGKHALDAVYLPFG KSAVVAKSLISGCGCTKRYLVTDTCVRFATHDDVAHDIQLPTAAIIPAWPVLATLPS GPELSKANAASVARDIVAPGP"
complement(4049..4081)
/gene="Rv3450c"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
complement(4228..4231)
/note="possible RBS, GGAG, for Rv3450c"
4388..5131
/gene="Rv3451"
4388..5131
/gene="Rv3451"
/note="Rv3451, (MTCY13E12.04), len: 247. Function: probable cutinase, similar to several, contains cutinase, serine active site motif (PS00155). Alternative start possible at 3733. FASTA results: CUN2_MYCTU_Q50664 probable cutinase cy339.08c precursor (219 aa) opt: 565; E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2% identity in 211 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:49:09 ; Search time 366.19 Seconds
(without alignments)
42.142 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaagtgattcgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	12412	17 AAT33537	BCG deletion regio
2	16.4	91.1	1535	16 AAQ89200	Mycobacterium tube
3	16.4	91.1	1535	17 AAT33656	M. tuberculosis ce
4	16.4	91.1	1535	19 AAV18647	DNA for M. tubercu
5	16	88.9	2642	22 AAH99048	Human EST-derived
6	16	88.9	2753	21 AAC76804	Human ORFX ORF2359
7	16	88.9	3125	22 AAH17859	Human CDNA sequenc
8	16	88.9	3262	21 AAAL6623	Human secreted pro
9	15.4	85.6	10594	20 AAAL3377	Enterococcus faeca
10	14.8	82.2	639	18 AAT67421	H. pylori flagella
11	14.8	82.2	1731	18 AAT68027	H. pylori flagella

12	14.8	82.2	96109	22 AAF28548	Genomic fragment #
13	14.4	80.0	21	AAF98936	Human gene single
c 14	14.4	80.0	69	21 AAA48232	T. reesei xylanase
c 15	14.4	80.0	573	19 AAV36098	DNA sequence of th
c 16	14.4	80.0	596	21 AAA48219	T. reesei xylanase
c 17	14.4	80.0	624	22 AAH56886	P. patens lipid met
c 18	14.4	80.0	625	22 AAH50894	Lipid biosynthesis
c 19	14.4	80.0	787	21 AAC35047	Arabidopsis thalia
c 20	14.4	80.0	1516	21 AAC33367	Arabidopsis thalia
c 21	14.4	80.0	1669	21 AAC39469	Arabidopsis thalia
c 22	14.4	80.0	1682	21 AAC35565	Arabidopsis thalia
c 23	14.4	80.0	2361	19 AAV70952	DNA sequence of th
c 24	14	77.8	300	20 AAZ12931	Human gene express
c 25	14	77.8	491	21 AAA64724	Human ovarian carc
c 26	14	77.8	545	22 AAF93587	lung carcinoma CHN
c 27	14	77.8	841	22 AAH01727	Human CDNA clone (
c 28	14	77.8	853	22 AAH07512	Human CDNA clone (
c 29	14	77.8	1753	21 AAC43745	Zea mays DNA fragm
c 30	14	77.8	1959	21 AAA07582	Marigold beta-cycl
c 31	14	77.8	2152	21 AAC74845	Human ORFX ORF400
c 32	14	77.8	2667	22 AAH16538	Human CDNA sequenc
c 33	14	77.8	2890	22 AAF82205	Melon ethylene rec
c 34	14	77.8	4797	22 AAF82217	Melon ethylene rec
c 35	14	77.8	4140	10 AAN92735	Sheep PrP gene for
c 36	14	77.8	11126	18 AAV74482	Staphylococcus aur
c 37	13.8	76.7	32	20 AAX87640	PCR primer lco375-
c 38	13.8	76.7	32	22 AAF72869	Primer #8. Homo s
c 39	13.8	76.7	390	11 AAO05314	3' portion of sequ
c 40	13.8	76.7	401	21 AAC30364	Human secreted pro
c 41	13.8	76.7	642	22 AAH02160	Shigella flexneri
c 42	13.8	76.7	730	21 AAC34255	Arabidopsis thalia
c 43	13.8	76.7	806	22 AAH07001	Human CDNA clone (
c 44	13.8	76.7	828	22 AAH04867	Human CDNA clone (
c 45	13.8	76.7	866	21 AAA54169	Sequence encoding

ALIGNMENTS

RESULT 1
AAT33537
ID AAT33537 standard; DNA; 12412 BP.
XX
AC AAT33537;
XX
DT 15-FEB-1998 (first entry)
XX
DE BCG deletion region 3 and flanking sequences.
XX
KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
XX
OS Mycobacterium bovis strain BCG.
XX
FH Key Location/Qualifiers
FT misc_feature 1406..10673
FT /*tag=a
FT /note= "BCG delta 1 deletion region"
XX
PN WO9625519-Al.
XX
PD 22-AUG-1996.
XX
PF 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahairas GG, Stover CK;
XX
DR WPI; 1996-393419/39.
XX

PT Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 XX bacterial infection
 PS Example 1; Fig 3; 66pp; English.
 XX
 CC This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta3. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT33535 and AAT33536) have also been detected.
 CC Identification involved screening a BCG cosmid library with a
 CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.
 XX
 SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;

Query Match 100.0%; Score 18; DB 17; Length 12412;
 Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;
 Matches 18; Conservative 0;

Qy 1 acatcaaaagtgcgcg 18
 Db 468 acatcaaaagtgcgcg 485

RESULT 2
 AAQ89200
 ID AAQ89200 standard; DNA; 1535 BP.
 AC AAQ89200;
 XX
 DT 09-OCT-1995 (first entry)
 XX
 DE Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
 DE protein.
 XX
 KW Vaccine; tuberculosis; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1535
 FT /*tag= a
 XX
 PN W09506726-A.
 XX
 PD 09-MAR-1995.
 XX
 PF 01-SEP-1994; 94WO-US09863.
 XX
 PR 02-SEP-1993; 93US-0118442.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Riley LW;
 XX
 DR WPI; 1995-115442/15.
 DR P-PSDB; AAR71931.
 XX
 CC DNA encoding for cellular uptake of Mycobacterium tuberculosis -
 PT used to develop prods for vaccines, passive immunisation and
 PT diagnosis and cellular uptake of other materials
 XX
 PS Claim 2; Page 9-11; 46pp; English.

XX The isolated DNA molecule of the invention confers on M.
 CC tuberculosis an ability to enter cells and to survive within
 CC macrophages. It encodes a polypeptide having a mol. wt. of about 50-
 CC 55 kDa, pref. 52 kDa. The AA sequence represents a highly
 CC hydrophilic protein with a hydrophobic region at its carboxy
 CC terminus. It could be a secreted protein, a cytoplasmic protein, or
 CC a surface protein with its carboxy terminus attached to the outer
 CC membrane of the organism. The deduced AA sequence is in AAR71931.
 XX
 SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 16; Length 1535;
 Best Local Similarity 94.4%; Pred. No. 12;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgcgcg 18
 Db 841 acgtcaaaagtgcgcg 858

RESULT 3
 AAT33656
 ID AAT33656 standard; DNA; 1535 BP.
 AC AAT33656;
 XX
 DT 22-NOV-1996 (first entry)
 XX
 DE M. tuberculosis cellular uptake gene region.
 XX
 KW Cellular uptake; cell entry; macrophage; passive immunisation;
 KW vaccine; gene therapy; ds.
 XX
 OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
 XX
 FH Key Location/Qualifiers
 FT CDS 181..810
 FT /*tag= a
 FT /label= ORF-1
 FT /product= cell entry protein
 FT CDS 886..1535
 FT /*tag= b
 FT /label= ORF-2
 FT /product= macrophage survival protein
 XX
 PN W09626275-A1.
 XX
 PD 29-AUG-1996.
 XX
 PF 20-FEB-1996; 96WO-US02155.
 XX
 PR 22-FEB-1995; 95US-0392210.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Riley LW;
 XX
 DR WPI; 1996-425086/42.
 DR P-PSDB; AAW02301.
 XX
 CC DNA giving M. tuberculosis ability to enter mammalian cells -
 PT and/or survive within macrophage(s); useful in vaccines to protect
 PT mammals against Mycobacterium tuberculosis infection
 XX
 PS Claim 2; Page 41; 67pp; English.

CC A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an
 CC ability to enter mammalian cells and to survive within macrophages.
 CC The encoded protein sequence is given in AAW02301. The DNA was obtd.
 CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II
 CC vector and screening recombinant E. coli strains for HeLa cell-

CC invasive clones. The DNA includes 2 separate coding regions (see
 CC also AAT33657-58) coding for the cell entry (AAW02302) and macrophage
 CC survival (AAW02303) proteins. It can be used to produce the cellular
 CC uptake proteins used as vaccines or to facilitate uptake of other
 CC materials, e.g. therapeutic genes.
 XX
 SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 17; Length 1535;
 Best Local Similarity 94.4%; Pred. No. 12;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acatcaaaagtattcgcg 18
 || |||||
 DB 841 acgtcaaaagtattcgcg 858

RESULT 4
 AAV18647
 ID AAV18647 standard; DNA; 1535 BP.
 XX
 AC AAV18647;
 XX
 DT 03-JUL-1998 (first entry)
 XX
 DE DNA for M. tuberculosis cellular uptake protein fragment.
 XX
 KW Cellular uptake protein; vaccine; infection; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1534
 FT /*tag= a
 FT /note= "stop codon not given"
 XX
 PN W09805784-A1.
 XX
 PD 12-FEB-1998.
 XX
 PF 06-AUG-1997; 97WO-US13056.
 XX
 PR 07-AUG-1996; 96US-0689411.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Chong P, Riley LW;
 XX
 DR WPI; 1998-145620/13.
 DR P-PSDB; AAW47541.
 XX
 PT Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
 PT therapeutic agents e.g. antibiotics, also useful in vaccines
 XX
 PS Disclosure; Pages 9-10; 82pp; English.
 XX
 CC The present sequence encodes a Mycobacterium tuberculosis cellular
 CC uptake protein fragment, which confers on M. tuberculosis an
 CC ability to enter mammalian cells and to survive within macrophages.
 CC The protein can be used in a vaccine to prevent M. tuberculosis
 CC infection, and provide for the uptake in cells of, e.g.
 CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
 CC raised against it can be used to treat mammals already exposed to
 CC M. tuberculosis, to induce a passive immunity and prevent disease
 CC occurrence.
 XX
 SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 19; Length 1535;
 Best Local Similarity 94.4%; Pred. No. 12;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acatcaaaagtattcgcg 18
 || |||||
 DB 841 acgtcaaaagtattcgcg 858

RESULT 5
 AAH99048
 ID AAH99048 standard; cDNA; 2642 BP.
 XX
 AC AAH99048;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 905.
 XX
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX
 DR WPI; 2001-476164/51.
 DR P-PSDB; AAM24389.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1; Page 724-724; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 SQ Sequence 2642 BP; 638 A; 703 C; 725 G; 576 T; 0 other;

Query Match 88.9%; Score 16; DB 22; Length 2642;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catcaaaagtattcgc 17
 |||||
 DB 642 catcaaaagtattcgc 657

RESULT 6
 AAC76804
 ID AAC76804 standard; cDNA; 2753 BP.
 XX

AC AAC76804;
 XX 08-FEB-2001 (first entry)
 XX Human ORFX ORF2359 polynucleotide sequence SEQ ID NO:4717.
 DE
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 XX 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 XX
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Leach M;
 XX WPI; 2000-602362/57.
 DR P-PSDB; AAB42595.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 5; Page 3901-3903; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB44397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 XX Sequence 2753 BP; 652 A; 746 C; 735 G; 619 T; 1 other;
 SQ

Query Match 88.9%; Score 16; DB 21; Length 2753;

Best Local Similarity 100.0%; Pred. No. 22;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 catcaaaagtgaattc 17
 Db 453 catcaaaagtgaattc 468
 RESULT 7
 AAH17859
 ID AAH17859 standard; cDNA; 3125 BP.
 XX AAH17859;
 AC
 XX 26-JUN-2001 (first entry)
 DT
 XX Human cDNA sequence SEQ ID NO:17563.
 DE
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 XX Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 XX Claim 8; SEQ ID 17564; 2547pp; 170 Kbp; English.
 PS
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dp primer; and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 3125 BP; 760 A; 797 C; 848 G; 720 T; 0 other;
 SQ

Query Match 88.9%; Score 16; DB 22; Length 3125;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaagtgatgcg 17
 |||||
 Db 856 catcaagtgatgcg 871

RESULT 8
 AAA16623
 ID AAA16623 standard; cDNA; 3262 BP.
 XX
 AC AAA16623;
 XX
 DT 16-JUN-2000 (first entry)
 XX
 DE Human secreted protein clone pt332_1 nucleotide sequence SEQ ID NO:11.
 XX
 KW Human; secreted protein; immunostimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antitachytic; antirheumatic; antiprotozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 XX
 OS Homo sapiens.
 XX
 XX WC200009552-A1.
 PN
 XX
 XX 24-FEB-2000.
 PD
 XX
 XX 13-AUG-1999; 99WO-US18298.
 PF
 XX
 XX 14-AUG-1998; 98US-0096622.
 PR
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX
 XX (GEMY) GENETICS INST INC.
 XX
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merbery D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX
 XX WPI: 2000-205979/18.
 DR P-PSDB; AAY94903.
 DR
 XX
 XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity -
 XX
 XX Claim 20; Page 474-475; 641pp; English.
 PS
 XX
 XX AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans

and animals. The polynucleotides can be used as markers for tissues in
 which the protein is preferentially expressed, as molecular weight
 markers on Southern gels, and as chromosome markers or tags to identify
 chromosomes or to map gene positions. The proteins can be used in the
 treatment of immune deficiencies and disorders, such as severe combined
 immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 infections. These infectious include human immunodeficiency virus (HIV),
 hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 candidiasis. The proteins can be used to treat autoimmune disorders such
 as connective tissue disease, multiple sclerosis, systemic lupus
 erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
 diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 autoimmune inflammatory eye disease. The proteins can also be used to
 treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 probes for the human secreted proteins from the present invention.

Sequence 3262 BP; 794 A; 859 C; 894 G; 715 T; 0 other;

Query Match 88.9%; Score 16; DB 21; Length 3262;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaagtgatgcg 17
 |||||
 Db 978 catcaagtgatgcg 993

RESULT 9
 AAA13377
 ID AAA13377 standard; tDNA; 10594 BP.
 XX
 AC AAA13377;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome contig SEQ ID NO:440.
 XX
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 XX
 OS Enterococcus faecalis.
 XX
 XX W000000000, A2.
 XX
 XX 12-MAY-1998
 XX
 XX 04-MAY-1998; 98WO-US08988.
 XX
 XX 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044041.
 PR 16-MAY-1997; 97US-0046655.
 XX
 XX (HUMAN) HUMAN GENOME SCI INC.
 XX
 XX Harash S, Hilton PJ, Kunsch CA;
 XX
 XX WPI: 1999-045171/04.
 XX
 XX New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX
 XX Claim 1; Page 1667-1672; 2084pp; English.
 PS
 XX A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAA12938 to AAA13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.

XX Sequence 10594 BP; 3197 A; 1733 C; 2323 G; 3337 T; 4 other;

Query Match 85.6%; Score 15.4; DB 20; Length 10594;
 Best Local Similarity 94.1%; Pred. No. 57;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaatgattcgc 17

|||||

Db 7892 acatcaaaatgattcgc 7908

RESULT 10

AAT67421
 ID AAT67421 standard; DNA; 639 BP.

XX AC AAT67421;

XX 08-JUL-1997 (first entry)

DE H. pylori flagella-associated protein ORF 19557055.aa.

XX Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW flagellum; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT CDS 1..639

FT /*tag= a

FT /transl_except= (pos:505..507, aa:Xaa)

FT /transl_except= (pos:514..516, aa:Xaa)

FT /transl_except= (pos:523..525, aa:Xaa)

FT /transl_except= (pos:526..528, aa:Xaa)

FT /transl_except= (pos:535..537, aa:Xaa)

FT /transl_except= (pos:544..546, aa:Xaa)

FT /transl_except= (pos:550..552, aa:Xaa)

FT /note= "Xaa = unknown; no stop codon given"

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

XX P-PSDB; AAW20191.

XX Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

XX Claim 1; Pages 163-164; 1481pp; English.

XX The present sequence encodes a Helicobacter pylori flagella-

CC associated protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX Sequence 639 BP; 222 A; 122 C; 142 G; 154 T; 9 other;

Query Match 82.2%; Score 14.8; DB 18; Length 639;
 Best Local Similarity 88.9%; Pred. No. 82;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 acatcaaaatgattcgc 18

|||||

Db 110 acatcaaaatgattcgc 127

RESULT 11

AAT68027

ID AAT68027 standard; DNA; 1731 BP.

XX AC AAT68027;

XX 15-JUL-1997 (first entry)

XX H. pylori flagella-associated protein ORF 07ge20415orf27.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT CDS 1..1731

FT /*tag= a

FT /note= "no stop codon given"

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

XX P-PSDB; AAW20774.

XX Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

XX Claim 1; Page 854; 1481pp; English.

XX This sequence encodes a H. pylori flagella-associated protein.

XX The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 1731 BP; 600 A; 302 C; 384 G; 445 T; 0 other;

Query Match 82.2%; Score 14.8; DB 18; Length 1731;
 Best Local Similarity 88.9%; Pred. No. 95;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 acatcaaaagtattcgcg 18
 ||||| ||||| ||||| |||||

Db 422 acatcaaaatcattcgcg 439

RESULT 12

AAF28548
 ID AAF28548 standard; DNA; 96109 BP.

XX
 AC AAF28548;

DT 04-APR-2001 (first entry)

XX Genomic fragment #35.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

OS WO200078968-A2.

PN 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -

PS Claim 1; Page 345-368; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

XX Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Query Match 82.2%; Score 14.8; DB 22; Length 96109;
 Best Local Similarity 88.9%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 acatcaaaatgattcgcg 18

||||| ||||| ||||| |||||

Db 14162 acatcaaaatgattcgcg 14179

RESULT 13

AAF96936

ID AAF96936 standard; DNA; 21 BP.

XX
 AC AAF96936;

DT 06-JUN-2001 (first entry)

XX Human gene single nucleotide polymorphism #1697.

XX Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
 KW polymorphism; vascular disease; coronary artery disease; forsenics;
 KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
 KW pulmonary embolism; paternity test; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Variation replace(11,G)

FT /tag- a

FT /standard_name- "single nucleotide polymorphism"

XX WO200118250-A2.

XX 15-MAR-2001.

PF 07-SEP-2000; 2000WO-US24503.

XX 10-SEP-1999; 99US-0153357.

PR 26-JUL-2000; 2000US-0220947.

PR 16-AUG-2000; 2000US-0225724.

XX (WIRED) WHITEHEAD INST BIOMEDICAL RES.

PA (MILL-) MILLENNIUM PHARM INC.

XX Lander ES, Garfield M, Ireland JS, Holk S, Daley GO, McCarthy JJ;

XX WPI; 2001-226749/04.

XX Nucleic acids comprising single nucleotide polymorphisms, useful in
 PT applications such as forensics, paternity testing, medicine, genetic
 PT analysis and phenotype correlations to diseases such as diabetes and
 PT atherosclerosis -

XX Examples; Page 161; 242pp; English.

XX The present invention provides a method of diagnosing a vascular disease
 CC in an individual, involving determining the sequence at various
 CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4
 CC genes. The sequences at a number of polymorphic sites are also provided
 CC in the specification. In particular, the method can be used in the
 CC diagnosis of atherosclerosis, myocardial infarction, coronary heart
 CC disease, stroke, peripheral vascular diseases, venous thromboembolism
 CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also
 CC useful in forensics, paternity testing, genetic analysis and phenotype
 CC correlations to diseases. The present sequence is an example of one of
 CC the human gene SNPs shown in the specification.

XX Sequence 21 BP; 7 A; 3 C; 6 G; 5 T; 0 other;

Query Match 80.0%; Score 14.4; DB 22; Length 21;

Best Local Similarity 93.8%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17
||| ||||| ||||| ||
Db 1 catcaaaagtattgcgc 16

RESULT 14
AAA48232/c
ID AAA48232 standard; cDNA; 69 BP.

XX
AC AAA48232;

XX
DT 28-SEP-2000 (first entry)

XX
DE T. reesei xylanase, xYtv-104, oligonucleotide.

XX
KW Xylanase; animal feed; digestion efficiency; thermostable;

KW feed pelleting; enzyme; xYtv-101; xYtv-102; Trx-103; xYtv-104;

KW xYtv-105; xYtv-106; xYtv-107; Trx-108; xYtv-109; xYtv-110; Trx-1;

KW xYtv-2; Trx-3; Trx-6; Trx-8; xYtv-4; xYtv-5; xYtv-7; xYn A; xYn B;

XX xYn; xYn C; xYn I; xYn II; ss.

XX
OS Trichoderma reesei.

XX
PN WO200029587-A1.

XX
PD 25-MAY-2000.

XX
PF 16-NOV-1999; 99WO-CA01093.

XX
PR 16-NOV-1998; 98US-0108504.

XX
PA (IOGE-) IOGEN CORP.

XX
PI Sung WL, Tolan JS;

XX
PI WPI; 2000-387799/33.

XX
PT Thermostable xylanases useful for preparing animal feeds especially
PT poultry or swine feed, exhibits optimal activity under physiological
PT conditions -

XX
PS Example 1; Page 74; 86pp; English.

XX
CC Xylanase enzymes are added to animal feeds to increase the efficiency of
CC digestion and assimilation of nutrients. Xylanases are preferentially
CC added during the feed pelleting process. To survive the pelleting
CC process and to have optimum activity in the animal, the xylanase needs to
CC have high thermostability, with optimum activity at physiological pH and
CC temperature. Various xylanases have some properties suitable for feed
CC applications but they lack the thermostability required to survive
CC food pelleting. The present sequence is oligonucleotide, xYtv-104, from
CC Trichoderma reesei. This sequence was used along with 17 other
CC overlapping oligonucleotides to construct the coding sequence for

CC Trichoderma reesei xylanase Trx. The other oligonucleotides were:

CC xYtv-101 (AAA48229), xYtv-102 (AAA48230), Trx-103 (AAA48231), xYtv-105

CC (AAA48233), xYtv-106 (AAA48234), xYtv-107 (AAA48235), Trx-108 (AAA48236),

CC xYtv-109 (AAA48223), xYtv-110 (AAA48237), Trx-1 (AAA48238), xYtv-2

CC (AAA48239), Trx-3 (AAA48240), xYtv-4 (AAA48241), xYtv-5 (AAA48242), Trx-6

CC (AAA48243), xYtv-7 (AAA48244) and Trx-8 (AAA48245). The resulting Trx

CC sequence was mutated to produce mutant xylanases with enhanced

CC thermostability (AAV99683, AAV99684, AAV99685, AAV99686, AAV99735 and

CC AAV99736) which would be useful for animal feeds, especially poultry and

XX swine feed.

XX
SQ Sequence 69 BP; 19 A; 15 C; 17 G; 18 T; 0 other;

Query Match 80.0%; Score 14.4; DB 21; Length 69;

Best Local Similarity 93.8%; Pred. No. 1e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17
||| ||||| ||||| ||
Db 34 CATTAAGTATTGCG 19

RESULT 5

AAV36098/c
ID AAV36098 standard; DNA; 573 BP.

XX
AC AAV36098;

XX
DT 02-SEP-1998 (first entry)

XX
DE DNA sequence of the specification.

XX
KW Family 11 xylanase; improve; thermophilicity; alkalophilicity;

KW thermotolerance; bleach; wood pulp; processing; wheat; maize;

KW digestibility-improving animal feed additive; starch production;

XX mutant; ds.

XX
OS Synthetic.

XX
PN EP828002-A2.

XX
PD 11-MAR-1998.

XX
PF 05-SEP-1997; 97EP-0115412.

XX
PR 09-SEP-1996; 96US-0709912.

XX
PA (CANA) NAT RES COUNCIL CANADA.

XX
PA (NARE-) NAT RES COUNCIL.

XX
PI Ishikawa K, Sung WL, Yaguchi M;

XX
PI WPI; 1998-161100/15.

XX
PT Modified xylanase enzymes - useful for improving wood pulp

XX
PT bleaching, etc.

XX
PS Disclosure; Page 50; 84pp; English.

XX
CC The present sequence appears in the specification, which describes a
CC method for modifying a Family 11 xylanase to improve its thermophilicity,
CC alkalophilicity and/or thermostability. This method comprises

CC modification of amino acids 10, 27 or 29 of Trichoderma reesei xylanase,

CC or corresponding oligonucleotide sequences in the N-terminal region

CC replacement of one or more amino acid sequences in the N-terminal region

CC with corresponding aligned sequences from another Family 11 xylanase to

CC form a chimeric xylanase and/or upstream extension of the N terminus by

CC addition of up to 10 amino acids. The modified xylanases are useful for

CC improving the bleaching of wood pulp by treatment at 55-75 degrees

CC Celsius and pH 7.5-9.0 for 3-180 minutes. They might also be useful as

CC digestibility-improving animal feed additives. They might also be useful

CC in the processing of wheat or maize for starch production.

XX
SQ Sequence 573 BP; 163 A; 127 C; 136 G; 147 T; 0 other;

Query Match 80.0%; Score 14.4; DB 19; Length 573;

Best Local Similarity 93.8%; Pred. No. 1.4e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17

||| ||||| ||||| ||

Db 471 CATTAAGTATTGCG 456

Search completed: April 2, 2002, 22:49:16

Job time: 9216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:46:04 : Search time 2977.34 Seconds
(without alignments)
99.736 Million cell updates/sec

Title: US-09-785-904-4

Perfect score: 18
Sequence: 1 catgccgcgtattgctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vi:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rod:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rod:*

36: em_htg_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
C 1	18	100.0	650	6	AR096715	AR096715 Sequence
C 2	18	100.0	650	6	AR147696	AR147696 Sequence
C 3	18	100.0	1535	6	AR096713	AR096713 Sequence
C 4	18	100.0	1535	6	AR147694	AR147694 Sequence
C 5	18	100.0	1604	1	MBU35021	U35021 Mycobacteri
C 6	18	100.0	3453	1	MTU43540	U43540 Mycobacteri
C 7	18	100.0	9281	1	MBDR3S1	U35017 Mycobacteri
C 8	18	100.0	9764	1	AE006921	AE006921 Mycobacte
C 9	18	100.0	10019	1	AF041819	AF041819 Mycobacte
C 10	18	100.0	12412	6	186264	186264 Sequence 18
C 11	18	100.0	17783	1	AE007028	AE007028 Mycobacte
C 12	18	100.0	32437	1	MTCY336	295586 Mycobacteri
C 13	18	100.0	38380	1	MTCY251	274410 Mycobacteri
C 14	18	100.0	43401	1	MTV13E12	295390 Mycobacteri
C 15	16.4	91.1	9600	14	AF176573	AF176573 Hepatitis
C 16	16.4	91.1	10096	1	AF026197	AF026197 Xanthomon
C 17	16.4	91.1	19304	1	AB045311	AB045311 Xanthomon
C 18	16	88.9	11811	1	AE004334	AE004334 Vibrio ch
C 19	15.4	85.6	480	3	AF051144	AF051144 Mamestra
C 20	15.4	85.6	924	1	AF208062	AF208062 Neisseria
C 21	15.4	85.6	925	1	AF208061	AF208061 Neisseria
C 22	15.4	85.6	1779	6	A96206	A96206 Sequence 23
C 23	15.4	85.6	3712	1	NM065788	U65788 Neisseria m
C 24	15.4	85.6	4862	1	AF121135	AF121135 Neisseria
C 25	15.4	85.6	4883	1	AF313394	AF313394 Neisseria
C 26	15.4	85.6	5859	1	NGU14554	U14554 Neisseria g
C 27	15.4	85.6	5859	6	AR025350	AR025350 Sequence
C 28	15.4	85.6	5859	6	AR025351	AR025351 Sequence
C 29	15.4	85.6	5859	6	AR111460	AR111460 Sequence
C 30	15.4	85.6	5859	6	124768	124768 Sequence 1
C 31	15.4	85.6	5859	6	124769	124769 Sequence 7
C 32	15.4	85.6	83921	9	AP001255	AP001255 Homo sapi
C 33	15.4	85.6	97811	2	AC019980	AC019980 Drosophil
C 34	15.4	85.6	126253	8	AC019018	AC019018 Arabidops
C 35	15.4	85.6	153865	2	AC091965	AC091965 Homo sapi
C 36	15.4	85.6	163687	3	AC008200	AC008200 Drosophil
C 37	15.4	85.6	174637	2	AC027490	AC027490 Homo sapi
C 38	15.4	85.6	177127	3	AC008199	AC008199 Drosophil
C 39	15.4	85.6	23313	3	AE003741	AE003741 Drosophil
C 40	15.4	85.6	340000	9	AP001686	AP001686 Homo sapi
C 41	15	83.3	439	14	AB030969	AB030969 Hepatitis
C 42	15	83.3	742	6	E06698	E06698 non A non B
C 43	15	83.3	742	6	149752	149752 Sequence 18
C 44	15	83.3	742	6	163390	163390 Sequence 18
C 45	15	83.3	932	14	S46012	S46012 polyprotein

ALIGNMENTS

RESULT 1
AR096715/c
LOCUS AR096715 650 bp DNA
DEFINITION Sequence 5 from patent US 6008201.
ACCESSION AR096715
VERSION AR096715.1 GI:10025753
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 650)
AUTHORS Riley, L.W.
TITLE DNA molecule encoding for cellular uptake of mycobacterium tuberculosis and uses thereof
JOURNAL Patent: US 6008201-A 5 28-DEC-1999;
FEATURES Location/Qualifiers
source 1..650
BASE COUNT 123 a 258 c 185 g 84 t
ORIGIN

08-SEP-2000

```

Query Match      100.0%; Score 18; DB 6; Length 650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
    |||
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 2
ARI147696/c
LOCUS      ARI147696      650 bp      DNA
DEFINITION Sequence 5 from patent US 6224881.
ACCESSION  ARI147696
VERSION     ARI147696.1 GI:15111786
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 650)
AUTHORS    Riley,L.W. and Chong,P.
TITLE      DNA molecule fragments encoding for cellular uptake of
           Mycobacterium tuberculosis and uses thereof
JOURNAL    Patent: US 6224881-A 5 01-MAY-2001;
           Location/Qualifiers
FEATURES   source
           1..650
           /organism="unknown"
BASE COUNT 123 a 258 c 185 g 84 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
    |||
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3
AR096713/c
LOCUS      AR096713      1535 bp      DNA
DEFINITION Sequence 1 from patent US 6008201.
ACCESSION  AR096713
VERSION     AR096713.1 GI:10025749
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1535)
AUTHORS    Riley,L.W.
TITLE      DNA molecule encoding for cellular uptake of mycobacterium
           tuberculosis and uses thereof
JOURNAL    Patent: US 6008201-A 1 28-DEC-1999;
           Location/Qualifiers
FEATURES   source
           1..1535
           /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 1535;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
    |||
Db 1074 CATGCCGTCGTATTGCTG 1057

us-09-785-904-4.rge
```

```

RESULT 4
ARI147694/c
LOCUS      ARI147694      1535 bp      DNA
DEFINITION Sequence 1 from patent US 6224881.
ACCESSION  ARI147694
VERSION     ARI147694.1 GI:15111784
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1535)
AUTHORS    Riley,L.W. and Chong,P.
TITLE      DNA molecule fragments encoding for cellular uptake of
           Mycobacterium tuberculosis and uses thereof
JOURNAL    Patent: US 6224881-A 1 01-MAY-2001;
           Location/Qualifiers
FEATURES   source
           1..1535
           /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 1535;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
    |||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 5
MBU35021/c
LOCUS      MBU35021      1604 bp      DNA
DEFINITION Mycobacterium bovis BCG DNA flanking deletion region 3.
ACCESSION  U35021
VERSION     U35021.1 GI:1049243
KEYWORDS   .
SOURCE     Mycobacterium bovis BCG.
ORGANISM   Mycobacterium bovis
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
           Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE  1 (bases 1 to 1604)
AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE      Molecular analysis of genetic differences between Mycobacterium
           bovis BCG and virulent M. bovis
JOURNAL    J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE    96200095
REFERENCE  2 (bases 1 to 1604)
AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
           Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES   Location/Qualifiers
           source
           1..1604
           /organism="Mycobacterium bovis"
           /sub_species="BCG"
           /db_xref="taxon:1765"
misc_feature 1400
           /note="site of deletion breakpoint in this organism."
           Virulent M. bovis has additional 10 kbp at this site."
BASE COUNT 337 a 637 c 416 g 214 t
ORIGIN

Query Match      100.0%; Score 18; DB 1; Length 1604;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
    |||
Db 698 CATGCCGTCGTATTGCTG 681
```



```
RESULT 8
AE006921 9764 bp DNA BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 7 of 280 of the
DEFINITION complete genome.
ACCESSION AE006921 AE00516
VERSION AE006921.1 GI:13879142
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 9764)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 9764)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..9764
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
122..1414
/gene="MT0095"
/note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; similar to SP:P16431 GB:X17506
PID:41684 PID:882614 GB:000096; identified by sequence
similarity; putative"
1449..2123
/gene="MT0096"
1449..2123
/gene="MT0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44319.1"
/db_xref="GI:13879143"
/translation="MSVKHAPSRVRLROTRSTVVKGRGSLSWRRVKRTGDLGLAVMG
GREETAVKPTGTPQKDMTIVYVDAGPGRSVRSVEVAAPAELEATVADPRHRH
ELDGSCTVGNKIVKAKLVVGSKFSTKMLFGLPRTVSRVTALPNHIVCSHPGH
RWRWFESLPTLTRVTFDYHAAGATKNGLKLYEMTGFSAKNAAGIFATLAKUSDQ
YARGRA"
2280..2873
/gene="MT0098"
2280..2873
/gene="MT0098"
/note="similar to GP:145427; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAK44320.1"
/db_xref="GI:13879144"
/translation="MDQPNANIHYDALLDMVPLTQCQVLDVGGCDGLLAARLARI
PYVTAVDIDAPVLRRAQTRFANPIRWLHADIMTAELPNAGDVAVVSNALHHIEDTR
TALSRLUGLVTGGTLAVVTFVTPSLRNLGLWHLTSWVACGMANKVRKGRWEHSIPKWP
```

```
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS
```

```
PROTLHELRSIVRALLPGACIRRLLYGRVLVTRAPV"
3002..3772
/gene="MT0099"
3002..3772
/gene="MT0099"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNQNRINRWELITCGLGHHVYYPDDAALAANKRASK"GLGE
VMKLRGCDFAIGAGPOGKAPEDAPLIMRGKALRQAIIRALGVERIVRALVLAAAM
AVWFERGARGAIGATIDKDIPLVLRAGKFKVDQMTVTHALEKALAAKPTLALITGMILA
AYVLAQVGVGLMLKRMGRKYFAVVATSTIFLPLEVHOLAKGIITTRVVTFSINVAAY
VYILISKRLGVKGRKAYDVERKEQLDLERAAMLT"
3791..4162
/gene="MT0099.1"
3791..4162
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MKGIRVVSKVTSYSAFFAIFTPARRARKAAKAVNLIDQVACRRKPTDEG
PTLCQHCUPGSAARALPTAAASKQSRVPATHCRPCCAPGAASALITCAICAFASVVE
VRAPRAATESLQSKAAARESH"
4206..4973
/gene="MT0100"
4206..4973
/gene="MT0100"
/note="similar to GB:AL00912; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine
nucleosidase/S-adenosylhomocysteine nucleosidase,
putative"
/gene="MT0101"
/db_xref="GI:13879147"
/translation="MAVTGVGICAIPOELAYLRGVLDKARQGVAAQIIFPSGQLDAHR
VLAAGMGKVTNGITATLLADREGCTVFTGVAGGLDPELCIGDIVIDRKYHIDIF
GLHDERIRPTQNGHITPIHEPTRELGYFVDIAYIDRVKRLDGFILAIPLSTAAGGGR
QRIYVYGTILTDQNYLHETRNRLHHELGMAVEMPGGAQVACASDITLWLVTRAL
SLAAGADSCVDENRFGVHEVAASSARVLELLPVLTAAC"
5105..7490
/gene="MT0101"
5105..7490
/gene="MT0101"
/note="similar to SP:P45840 PID:559912; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTTAVTGHHSVORIOLRISGMSCSACAHKVESTLNKLPCVRA
AVNFGTRVATIDTSEAVIAALCOQVRAGYQADLCTDDGRSASIPDADHAKQLIKL
AIAAVLFYVADLSYMFQVVPATRTGQWVLSALALPVTWMAAPFHVRARNARHH
AASMETLISVGIATAATVNSLYTFGNHSPITERSGIQWALLGSDAIYFEVAGVTVFVL
VGRYPEAKAKSQAGSALKALAAKSAKEVALLPDGSEMVIPADEIKEQRFVVRPGQI
VAAGLAVDGSAAVDMSAMTGEAKPTRVRPGQVIGGTVTLDGRIIVEAAAQVADYOF
AGWRLVHQQAQKADARLADRISSVFVPVAVI AALTAAQWL IAGGUPHVASAAL
AVLITACPCALGATPTAMVVASGRGAQIGIFLKGYKSLEATRAVDTVVYDKTGTLTT
GRIQSVTAAPAGVNEADQVLAATAVAASEHSVALATAATTRDAVTDTRKAIIPGRG
VSGTSGRAVRVKGPSWIGSSCHPNMRAARHAEISLGETAVFVDEGPGCVIADAV
AVKDSADAVAADRGLRMTLLTGDNPESAATAVRVIGDEVIDIADIIPEGRKVDYIEQ
LHDGRHVVMYGDGINDCGPALRADLGNAGIGRTDVAICAADIIIVRHILDVVPLAD
LARATPRVVKLNWAFQNTAAIPVAAAGLLNPLVAGAAAFSFEFFVSNLSLRKRF
GRYPLGCGTVGGPQMTAIPSSA"
complement(7337..8185)
```

[illegible]

CDS

4277..4957
/gene="bioD"
/EC_number="6.3.3.3"
/note="BioD; similar to MTCY336.33c"
/codon_start=1
/transl_table=11
/product="putative dethiobiotin synthetase"
/protein_id="AAB96958.1"
/db_xref="GI:2791847"
/translation="MTLVVTGTGTGVTGKTVGCAALASARQAGIDVAVCKPVQGTGA
RGDDLEVERGLAGVQLAGLARYPOPMAPAAAEHAGMALPARDQIVRLIADLDRPG
RLTIVGAGGLLELAPGVTLRDVAVDVAVALVVTADLGVINHTKLTLEALAAQO
VSCAGLVIGSWPPPPGLVAASNRKSALARIATVRAALPAGAASLDAGDFAAMSAAFDR
NWAGLVG"
4957..5466
/note="ORF6; similar to MTCY336.32c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96959.1"
/db_xref="GI:2791848"
/translation="MVHSIELVFDSTDEAATRIIWAGLAAAGIPSOAPASRPVSLAV
AERTAPEDEPLGAVARRLPLDCVIGAPVLFGRANVFTRLVPTSELLAHAEVHRL
CGPHLAPAMANSIPGQWTAHVTLARRVGGHQLGRALRIAGRPISRIDRPAFLRWMDG
NTRAYLLIG"
complement(5612..6478)
/note="ORF7; similar to MTCY251.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96960.1"
/db_xref="GI:2791849"
/translation="MDNLHPGDLTDTERARKGITLSNOQYDCMSLGSVLTPOARA
TFEAYLAKLAPGATNPDDHTPVIDITPDAAIDRDTPSOQRNIDGLIAGLALIAS
GELGQHGILPVSIVVTTTLDTQAGAGFTGGGTLPLMADVI RMTSHAHYSPASR
YQOALFDHGTPLALYHTRKLASPAQRIMLFANDRGCTKPGCDAPAYHSAHHVGTWS
TGRTDITDITLACDPDLAEKGTWTRKNTGHTEWLPPPHLDHCQRTNTFFHHEKL
LRHNDENHDDP"
complement(6339..7007)
/note="ORF8; similar to MTCY336.16"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96961.1"
/db_xref="GI:2791850"
/translation="MLANSRELVFEDALDAELDRDLSFVEVITPERLRSLELRE
CLVRLPVGHTLIINOLDTQASEELGTLCCALANRLITKPDAAALRTADAADLGP
RALTEPLAQPOLATAQOGLIGEAAHKVIRALFRPARRGGCVHPPGRSRSGHQ
SRSSSRAGPLRPAGHGATPPRRPHRIHTNPQTRHHHPQPAHKHVTAKWLPDHPH
AGHL"
7455..8504
/gene="bioB"
7455..8504
/gene="bioB"
/note="BioB; similar to MTCY336.15c"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAB96962.1"
/db_xref="GI:2791851"
/translation="MTQAATRPTNDAGODGNNSDILVVAQVQLQRGGLNODQVLA
VLOLPDRLEELALAEHVRMCQGEVEVGIISLKGCCPEDCHFCOSGLFASPV
RSALDIPSLVEAAKQATKSGATEFCIVAARVGRFDMQAQVAEGEAIARNEVEINIA
CSLGLMTAEQVDQLAARGVYHNHLETSRFSFANVVTTHWEERWOTLSMYKRDAGME
VCCGGILGHEGTQQRAEFAELAEIGDPDEPLNPLNPRGTFPADLEVPVPGDALK
VAAPRALPTMLRFPAGGREITLGLDCAKRGILGGINAVIGNVLYTLTLGRPAAEDLEL
LDEQLQPLKALNASL"
8825..9406
/note="ORF10; similar to MTCY336.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96963.1"

CDS

1456 a 4134 c 4655 q 1774 L
ORIGIN

Query Match 100.0%; Score 18; DB 1; Length 10019;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtcgattgctg 18
|||||

Db 6386 CATGCCGCTGATTGCTG 6403

RESULT 10
186264/c
LOCUS 186264 12412 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 18 from patent US 5700683.
ACCESSION 186264
VERSION T86264.1 GI:3205982
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12412)
AUTHORS Slover, C. Kendall and Mahairas, G.G.
TITLE Virulence-attenuating genetic deletions deleted from mycobacterium
H3G
JOURNAL Patent: US 5700683-A 18 23-DEC-1997;
FEATURES
source location/Qualifiers
1..12412
/organism="unknown"
BASE COUNT 2146 a 4274 c 3875 g 1917 t 200 others
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 12412;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtcgattgctg 18
|||||

Db 705 CATGCCGCTGATTGCTG 688

RESULT 11
AE007028
LOCUS 17783 bp DNA BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 114 of the
complete genome.
ACCESSION AE007028 AE000516
VERSION AE007028.1 GI:13881250
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 17783)

AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Winn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 17783)
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Winn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Direct Submission
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..17783 /organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" complement(104..826) /gene="MT1622.1" complement(104..826) /gene="MT1622.1" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45890.1" /db_xref="GI:13881251" /translation="MLAKLAAGATNPDDHTPTVIDTTPDAAAIDRTRSOAQRNHDGL LAGRLATSKLQGHNGLPVSIYVTTTLDLQFGACKGKTGGTLLPMADYVIRPMSH AHYSPASGRYPQAFIDHGTPTLALYHTRKASPAQRIMLFANDRGCKPKCCDAPAYHS QAHVVTATGRTGRTDIDTLACDPDRLAELKGTWTKHTNGHTWELPPPHLDHGQPR TNTFHHEKLLRHNDHNDP" complement(831..1499) /gene="MT1623" complement(831..1499) /gene="MT1623" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45891.1" /db_xref="GI:13881252" /translation="MLANSRELVVEFDALDAELDKLDEVSFEVLTTPKLSLEKLE CIVRLPVGHTLMLNDTSRELEAGTLCALANKLRITKPMALIKRIADAAULGPK RALTEPLAGPITATATAGRGLICEAHIKVIRALFPAKRRGGLVHHPGRKSRKQHQ SSISRRAGPLRPAGHGLATPRKRPHIRTPQTRHHPQATIRHIVTAKRIIDPTIS AGHL" 1947..2996 /gene="MT1624" 1947..2996 /gene="MT1624" /note="similar to GP:1680650; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="biotin synthase" /protein_id="AAK45892.1" /db_xref="GI:13881253" /translation="MTQAATRTPDNDAGDGGNSDILVVAHQOVLORGECSGLFASPV VLOLPDDRLEALLAHVRWRCGPEVEGIIISLKTGCGPCDPCSCSGLFASPV RSAMLDISLVEAAKOTAKATCFIVAAVRGPDRLMAOVAAGIEAIRNEVEINIA CSGLMLTAEOVDLAQRGVHYNNILETABSFEANVTHTWEERWOTLSMRDAGME VCCGILTGMEQLQORAEYFAELAEELGDEVPFLNLPNRCPTPPADLEVHPVGDALKA VAAFRALPRTMLRPAAGREITLGDUGAKRGLGINAVIGNVLTTLGRPAEDLEL LDELQMLKALNASI" 2997..3236 /gene="MT1625"
gene	1947..2996 /gene="MT1624" 1947..2996 /gene="MT1624" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45893.1" /db_xref="GI:13881254" /translation="MPDGHFEGSLMVEFGNIAGATGAEWIGRPPHEELQRKVKPLPLPS HFFYFPAGYGHAVGTIVLRSDKDELAFMGILIPQVATATQLLYRTTMYNGMPEATVTT VIVPAELAHQVETLLSYCAIDAMSSRCFPFSYALRRKAKALASITPMELIMIISAAAL EIRAVYSPHJEGFKELWASVYFPGYKVLGDLRAALNSRCKVLSPATIGLWYSGGGL A:AAAEANGLVATDILIVGAVIGSVCJLITFKRLMCTLLALPALVVAALQIISVVE L:ARVLEKIANHGHGCHLELLEMTETIUDVATMAGKUNGDFLDEPLEHLLSTPFSHVS L:PKLSASVFPFVPLVIVGAVHDYLIIDVSDIALDALAGTAGGANTVTHRDLFSGHVS HLLSAPMTLRWLTDRFAGKPLTDHRVTRTWTPTIFNUMTYAGMAKLAIVAAKVLTRKTL SKKPL" complement(5660..6193) /gene="MT1629" complement(5660..6193) /gene="MT1629" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45897.1" /db_xref="GI:13881258" /translation="MTSSVRKQLAKFKVDLRELAHLLEQLAVFSDPHRLPGIRMIASVY GVVSPATPELPADTRWHPVSLSPPMADHGPMVTHARTLIAKMSYTNIGFALAPKE FALSTLDITGAALGYQVDATNLQVLRARRVITQTGTIAQSGRSGRPAALYRFTDSH QLRVTFDEAALRPPQGL" 6419..7468 /gene="MT1630" 6419..7468 /gene="MT1630" /note="similar to SP:P11458 GB:X12713 PID:581137 GH:00000969
CDS	2997..3236 /gene="MT1625" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45893.1" /db_xref="GI:13881254" /translation="MVKILVAGKQAPVAAAGVYNYVYIGELADTATPTAARMGLPEPPKPC AOCGRMVDVQVRPDGMWAKSRHQVDSADLATOR" 3317..3898 /gene="MT1626" 3317..3898 /gene="MT1626" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45894.1" /db_xref="GI:13881255" /translation="MIGLSATGVLGVGLWAMIAPIIHAVVAITKAGRPVHFYILGSESO NFFTAIPMLGLLSVLAIVASALAMKWRHFGPOMVAGISLIGLITAAAIANGVCAIVV RIKYGALDFDTVPVLSKGNIAITYYTQAPPVFFARPLQIALTIMWPAGIASIVYALLA AGTAKDIDLGYPAPVDPNSNAITEALETPQAPVS" complement(3953..4054) /gene="MT1627" complement(3953..4054) /gene="MT1627" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45895.1" /db_xref="GI:13881256" /translation="MRSATHTCANSDLAHLNLTVDLINOQFDLPLESKR" complement(4063..5430) /gene="MT1628" complement(4063..5430) /gene="MT1628" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45896.1" /db_xref="GI:13881257" /translation="MPDGHFEGSLMVEFGNIAGATGAEWIGRPPHEELQRKVKPLPLPS HFFYFPAGYGHAVGTIVLRSDKDELAFMGILIPQVATATQLLYRTTMYNGMPEATVTT VIVPAELAHQVETLLSYCAIDAMSSRCFPFSYALRRKAKALASITPMELIMIISAAAL EIRAVYSPHJEGFKELWASVYFPGYKVLGDLRAALNSRCKVLSPATIGLWYSGGGL A:AAAEANGLVATDILIVGAVIGSVCJLITFKRLMCTLLALPALVVAALQIISVVE L:ARVLEKIANHGHGCHLELLEMTETIUDVATMAGKUNGDFLDEPLEHLLSTPFSHVS L:PKLSASVFPFVPLVIVGAVHDYLIIDVSDIALDALAGTAGGANTVTHRDLFSGHVS HLLSAPMTLRWLTDRFAGKPLTDHRVTRTWTPTIFNUMTYAGMAKLAIVAAKVLTRKTL SKKPL" complement(5660..6193) /gene="MT1629" complement(5660..6193) /gene="MT1629" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45897.1" /db_xref="GI:13881258" /translation="MTSSVRKQLAKFKVDLRELAHLLEQLAVFSDPHRLPGIRMIASVY GVVSPATPELPADTRWHPVSLSPPMADHGPMVTHARTLIAKMSYTNIGFALAPKE FALSTLDITGAALGYQVDATNLQVLRARRVITQTGTIAQSGRSGRPAALYRFTDSH QLRVTFDEAALRPPQGL" 6419..7468 /gene="MT1630" 6419..7468 /gene="MT1630" /note="similar to SP:P11458 GB:X12713 PID:581137 GH:00000969

PID:1651334; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="quinolinate synthetase"

/protein_id="AAK45898.1"

/db_xref="GI:13881259"

/translation="MTVNLNRDTFLVDELTAADITNTPLGVGGVGDGKRWAAEIRRLAHL
RGATVLAHNTQLPAIQADVADHVGDSIALSRVAAEAPEDTIVFCGVHFMMAETAKILSPH
KTVLIPQKAGCSLADSIPTDELRAWKDEHPGVAVVSYNTTAANKALTDICTSSNA
VDVVASIDPQREVLPQDFQFLGAHVRVTGKLNHYWAGBCHVHAGINGDELADQARA
HPDELFPVHPECCGATSAIYLAGEGAPFAERVKILSTGGMLEAAHTTRAROVLAATEV
GMLHOLRAAPEVDFAVRANDRASCKYMKMTTPALLRLCLIVEGADEVIVDPGTAASCKR
SVORMIEIGHPGGGE"

7468..9051

/gene="MT1631"

7468..9051

/gene="MT1631"

/note="similar to GP:608530; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="L-aspartate oxidase"

/protein_id="AAK45899.1"

/db_xref="GI:13881260"

/translation="MAGPAWRDAADVIVIGTVAGLAALAAADPAGHSVVVLKAAQT
HVTATHYAQGGIAVVLDPNDSDVDHADVADTLAAGAGLCDDPAVYSIWAICYRAVTDLV
CAGARLDESVPGRWALTREGHSRRIRVHAGGDATGAEVORLQDAAGMLDIETGHVA
LRVLHDGTFTGLLVVRPDGCGIISAPSVILATGGLGHLYSATNPAGSTGDTALGGL
WAGVASDLETFIQHPTMLFAGRAGSRRLPTAIEIRGEAILVDROGNSITAGVHPMG
DLAPRDVVAAGIDARLAKATGDCPVYLDQAGEFPASRFPPTVATSCRAAGIDVPKQPI
VYRGAHYSCGGIVTDVYGOTELGLYAAGEVARTGLHGANRLASNLLEGLVGGGRAG
KAAAHAAAGRSRSTSSATWPEPISVLTDRGDLQRAMSRDASMYKAAAGLHRLGDS
LSGADVQVACRRQFEDVALILVAQSVTAALATETSRGCHHRAEYPCCTVPEQARSIV
VRGADANAVCQALVAVC"

9021..9908

/gene="MT1632"

9021..9908

/gene="MT1632"

/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262
PID:456041; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

Query Match 100.0%; Score 18; DB 1; Length 17783;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgctgctgtattgctg 18
|||||

Db 878 CATGCCGTCGATTGCTG 895

RESULT 12

MTCY336

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis H37Rv

Mycobacterium tuberculosis H37Rv

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Mycobacteriaceae;

Mycobacterium; Mycobacterium tuberculosis complex.

1 (bases 1 to 32437)

Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
Tekaita,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,K., Devlin,K., Feltwell,I., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squires,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and

Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)

98295987

2 (bases 1 to 32437)

Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
on Jun 27, 1998 this sequence version replaced gi:211723.

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

location/Qualifiers

1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"

1..274
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"

33..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"

/clone="Y336"
complement(39..222H)
/gene="RV1565c"
complement(39..222H)
/note="RV1565c", (MTCY336, 48), len: 729. Unknown membrane
protein, some similarity to O05402 HYPOTHEICAL 72.2 KD
PROTEIN from B. subtilis (634 aa) opt: 384 E(); 4-8e-17;
29.1% identity in 478 aa overlap and to hypothetical
protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, Y392.HAEIN P43993 hi0392 (245 aa) opt: 265;
E(): 5.5e-10; 28.3% identity in 247 aa overlap. Tbpase
score is 0.930"

/codon_start=1

/transl_table=11

/product="hypothetical protein RV1565c"

/protein_id="CA809072.1"

/db_xref="GI:2117272"

/db_xref="SPTREMBL:O06625"

/translation="MLTISPPRPALTPPEALPPVTMGTRTTGFRHIDLKLRGVAIA
LVVPHVWFGVSGGVDFLALSGFEGKILRAALNDLSLSPAEVIRKILRLPPA
HTWSVGQGYFLAPLLVAGCAVLLRLFRGRAPYLRPMVRLTAVATAALAIISGAL
VAHAYQATAYNTFAWELLAGAVGVPHVPMVRLTAVATAALAIISGAL
IDGVKEFFGPWALVPVATMLLAGANROGHPGTDRPLNPLATAPILVAGMA
YSWLHWPPLLIWLSYTGHRHAFVEAGVALLVSGLLAYLTRIVEDPLRYAPAGV
RSPAAYPIPHRLRLRPTIVLGSVALLGVALTATSTFTRHIVTVOAAKLELSGLS
SHDYPGARALIDHVRVKPLKMRPTVLEVRHDLPTSKDGCISDFVNPALINCTYGDVD
APRTTALAGGSHAEHMLTALDLLGRMHFKVVTYLLKMGCPSTEEVPIIMGNAPYPO

TITLE
JOURNAL

COMMENT
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
(n Jun 27, 1998 this sequence version replaced gi:1405752.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES
source
Location/Qualifiers
1..38380
/organism="Mycobacterium tuberculosis H37Rv"
/db_xref="H37Rv"
/db_xref="taxon:83332"
<1..35889
/organism="Mycobacterium tuberculosis H47Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="H251"
6..956
/gene="hycB"
6..956
/gene="hycD"
/note="Rv0084, (MTCY251.02), len: 316, hycB, similar to HYCD_ECOLI P16430 formate hydrogenlyase subunit 4 (30 aa) FASTA scores: opt: 570, z-score: 591.6, E(): 2.1e-26, (33.8% identity in 305 aa overlap) and NUOH_ECOLI P33603nadh dehydrogenase i chain h (325 aa) FASTA scores, opt: 207, z-score: 220.7, E(): 9.5e-06, (26.5% identity in 260 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycB"
/protein_id="CAA98920.1"
/db_xref="GI:1405754"
/db_xref="SPTREMBL:Q10881"
/translation="MSYLAGAAQIGGVVGAPIVIGMTRQVRAKRWKRAKAGILIQWRDLKQLGKQITPAGTTIVFAAAPVIVAGTTLTLLIATPVLATGSPIDPSADLPFAVVGLLFGTVALTGADTGTSCGMSGASREITIAALVEPTILLAFALSIPAGSANLUGALVASTIDHPGHVWSLAVLAFVAVIVIVAEETGRUPVDNPAPHILEITVWRHAMVLEYAGPRALVSWAAGMRLTLLALLANLFLPWGIAGAAPTALDVLITGVVAVAAKVALILAVILATFEVFLAKLRVPELLAGSFLALLAVTAANFFTVGA"
967..1629
/gene="hycP"
967..1629
/gene="hycP"
/note="Rv0085, (MTCY251.03), len: 220, hycP, similar to spiP77524|HYFE_ECOLI HYDROGENASE-4 COMPONENT F (216 aa).FASTA scores: opt: 204 z-score: 272.1 E():1.2e-07; 25.5% identity in 216 aa overlap"
/codon_start=1
/transl_table=11
/product="hycP"
/protein_id="CAA98921.1"
/db_xref="GI:1405755"
/translation="MSNANFSLIVDFAGGGLVLASVLTVMRRDLRAIVRLAWQGAALAAIPLLGIRNDRLIATVGLAVIALVLPWILARAVGAEEAAQREAPPIVNTASSLLITAGLTLTAFATITQPVNLEPGVTINAVPAAPAVVLIATFVMTTKIHAVSQAAQFL

gene
CDS

gene
CDS

gene
CDS

MLDNGTAATAFLLTAGVPLIVELGASLDVLFAVIVIGVLTGRLRLRIFGDAIILDKI.REL
RU"
1629..3095
/gene="hycQ"
1629..3095
/gene="hycQ"
/note="Rv0086, (MTCY251.04), oxidoreductase, len: 488, hycQ, most similar to spiP77437|HYFE_ECOLI HYDROGENASE-4 COMPONENT F (EC 1.- (526 aa), FASTA scores: opt: 948 z-score: 1117.4 E(): 0; 35.9% identity in 493 aa overlap. Also similar to E. coli D9087711 & NUOH_ECOLI P33607nadh dehydrogenase i chain 1 (613 aa) FASTA scores, opt: 460, z-score: 354.9, E():1.2e-13, (27.9% identity in 488 aa overlap), and to NUOH_ECOLI P33608, nadh dehydrogenase i chain n (425 aa). FASTA scores, opt: 375, z-score: 371.4, E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/translation="MTGLLAAIILAPLAASIASLITGWRRTTATLTALSATFVACAVAMPFWMGSGAGPGLGGLIKRAVLTVMVLVIGVGLTATAASIGYIITPELAIHGLIDGRSAKLYGVLTPEAFIAAMVIAVANNIVIAIVTATVITAFVGHRTKRTALFAFKYVVICSGYIAVAFICITVLIYPAARDSCAAAGALNLDILAEHAGLIDHICVAPKAGCLLLIGYAKAGLFFPHITLADAHISQAPVPSALMSGVLLAVAFSVILRLRLILLANSQPAYLRNGLAVGLATLLIIVAVLITVTGDKYKRLIAYSMEIMGLIATAAAATTIATIAALLILHVLAIIGTKTVEFLAGLQAQAHDSIAIDITGVMRSRRLIGVSPFVGLIVLGLPPFAMFASHLIAKSLANERLAWLGAALLLIATIGFTALAKNSGRMLIGTTPAAATATVTPATAAALMVGIVVSVAAIGITAGPLADLIGIAASNVGLI"
3092..4570
/gene="hycE"
3092..4570
/gene="hycE"
/note="Rv0087, (MTCY251.05), oxidoreductase, len: 492, hycE, similar to HYCE_ECOLI P16431 formate hydrogenlyase subunit 5 (569 aa) FASTA scores, opt: 680, z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa overlap) and to NUOH_ECOLI P33600nadh dehydrogenase i chain d, (407 aa) FASTA scores, opt: 245, z-score: 293.1, E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/translation="MILHDLGGLDGLHRLKRYV
VYFLAVYPRKRVHAAVVAQNPITFSLAYLSPIAGREFEFEMADYVIRVAGHKPKKLVYAHRTVPHIRPKFLAGVADPETHCAPEFLAVEGVVEIIPAGVIAAGLIEGHIFKEVAGETIVRLKALIMFVIHGTETKLEFHGKPATAAVDLAEKISGITSAAUALHSLAIHIALGTELPHEVIRIKALIVELERLYNHAADIGALANIVGYSLANAHOAKREINLRRNAAVTGHRLIRGAIKAGCAVALRALPDTDELAALADIAEATVILANSVYRDFAGTAVHHDDASALGICLYVAKASGLRSIDRSEHPHTIVLPITETGAPICILARYIVKRDDEF AASAAALAHIVESHPTCIIEYAATLHPVGPSSGILVFGWGTIVHKVEIIVDKLITRKAKVDPSPENWHPALIVAMADITVDPDFPLANKSPNQSYAGNDI"
4605..5279
/gene="Rv0088"
4605..5279
/gene="Rv0088"
/note="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/translation="MSYKIAPSRVRLRQRPSTVYKGRSGLSMRVYKTCILGLAVMG
GREYRAVKPGCTGTPQPKGMNMTVVVDAGPGRSVRSVEVAAPAAELFAIVADPVRHRELDGSGVGRVTKIPGPAKLVGSKFSTKMLFGDLBYRTTSRVTLKPLKPNLCWSNIIIGH
HWRWFFTSISPTLTRVTETFDYHAAGAATKNGLKPEYEMTGFAKSAAGATATATLAKISDQYAKRA"

```
gene      5436..6029
/gene="Rv0089"
CDS       5436..6029
/gene="Rv0089"
/note="Rv0089, (MTCY251.07), len: 197. unknown, some
similarity to sp1p12999|BIOC_ECOLI BIOTIN SYNTHESIS
PROTEIN BIOC. (251 aa). FASTA scores: opt: 202 z-score:
262.0 E(): 4.5e-07; 39.0% identity in 118 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0089"
/protein_id="CAA98925.1"
/db_xref="GI:1405759"
/db_xref="SWISS-PROT:Q10886"
/translation="MDQPMNATHYDALIDAMVPLGTCQVLDVCGCGCLLAARLARH
PYTVADIDAPVLRRAQTRFANAPIKHLHADIMTAELPNAGFDVAVNSALHHEUDIR
TAUSRLGGLVTPGGTGLAVVTFTPSLRNLGWLHLSWVACCMANKVKCKWEHSAPIKWP
PPQTLHELRSVHRALLPGACIRRLLYGRVLVTWRAPV"
6147..6152
/note="aggag, potential rbs upstream of Rv0090"
/gene="Rv0090"
/gene="Rv0090"
/note="Rv0090"
/note="Rv0090, (MTCY251.08), len: 256. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0090"

Query Match      100.0%; Score 18; DB 1; Length 38380;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgcgtatgctg 18
      |||||
DB 12159 CATGCCGTCGATTGCTG 12176

RESULT 14
MTCY13E12/c
LOCUS      MTCY13E12      43401 bp      DNA      BCT      03-AUG-2001
DEFINITION      Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
ACCESSION      295390 AL123456
VERSION        295390.1 GI:3261766
KEYWORDS
SOURCE      Mycobacterium tuberculosis H37Rv.
ORGANISM      Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 43401)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,K., Devlin,K., Feltwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)
98295987
2 (bases 1 to 43401)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2104370.
```

```
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, atg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1..43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
<1..1481
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
/clone="Y77"
78..1481
/gene="Rv3448"
78..164
/feature
/feature="Rv3448"
/feature="PS00402 Binding-protein-dependent transport
systems:inner membrane comp signature"
78..1481
/feature="Rv3448"
/feature="MTCY77.20", len: 467 aa. Unknown membrane
protein, contains PS00402 Binding-protein-dependent
transport systems:inner membrane comp signature. Some
similarity to AL0219.40|MTV035_18 Mycobacterium
tuberculosis (472 aa) fasta score, opt: 429 z-score:
488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3448"
/protein_id="CA808736.1"
/db_xref="GI:3261767"
/db_xref="SPRKBML:033354"
/transl_table=11
/product="MPTSPHMLRRVTWAGAAQAVDITLPAAPVATVILPSIVDILGDR
GASPTAARYQISALGAPALPNATTLAQCGIRDGAVLVLRHSSAGPTTPRCDDVAEAV
AAALDITTRAPQDKETRELSALAAACITAGGGLMVRNALGCTNVTYSDTAGVAAA
GLAALFAVIAKRYRRLFAVLTLSVLTATFGVAGLVAVPVGIVSLVAHAAA
PSVLMKRTGCGITLTAVANCAVVAATVAVGAIITAAPVATIGSLATLASFGILEVS
ARMAVLGAGISRLPPALNDDADALTTLTRKADAWLTSLAFAASATIGA
IGTAVATHGIRHSSMGCIALAAVTCALLLLRARSADYRSLVFAICGLTTVATPTVA
ADRALEHGPWIAALTATLAAVAMFLGFPVAPALSPTVYRTIELLECLALAMVPLTA
WLCGAYSAVRHLDLTWT"
611..>43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
/clone="Y13E12"
1478..2845
/feature="Rv3449"
1478..2845
/feature="Rv3449"
/feature="Rv3449"
/feature="Rv3449, (MTCY13E12.02), len: 455. Function:
probable precursor of serine protease. Has putative signal
peptide N-terminus and hydrophobic stretch at
C-terminus. Contains three signatures typical of subtilase
family: aspartic acid active site (PS00136), histidine
active site (PS00137), serine active site (PS00138). FASTA
results: 053863 SERINEPROTEASE (390 aa) opt: 241; E():
8e-07; (38.0% identity in 387 aa overlap), similar to
MTCY15F10.29 (45.7% identity in 451 aa overlap)"
```

```

/codon_start=1
/product="hypothetical protein Rv3449"
/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/translation="MTTSRLRLVLSALATLSGLGTPVAHAYSPPIIDRWLPESAL
PRLPPVQREVCVEVTEAGRAFRASQAQLADLDVWRLTRGAQRVAVIDTGV
RRLPKVAGDVFTGDTGADCAHTLAVGTIAAAPDAOSDNFSGVAPDVTLSI
RQSSKRAPGDPSTGVGDVTHAKAVRTAADLGASVINISSTACVPAARAPDRAL
GAALYADFTLVAVAAAGTGAACFPQAQVTRDSTVTVSPAWYDDYVLTVCS
VNAQEPSTFLAGPDWDDVAATGAVTSLSFGDGTNRLLGGQHGSIPISTSTAAHP
VSLAALIRAFRPTLTARQVMRIESTAHHPAGMDPLVGNGTVDALAAVSSDSIPQA
GTATSDAPVAVPVPRRSTPGSPDRALHTAFAGAAICLLALMATLATSRRLRPRN
GIAGD"
misc_feature
1757..1789
/gene="Rv3449"
/note="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862..1894
/gene="Rv3449"
/note="PS00137 Serine proteases, subtilase family,
histidine active site"
2456..2488
/gene="Rv3449"
/note="PS00138 Serine proteases, subtilase family, serine
active site"
complement(2810..4222)
/gene="Rv3450c"
complement(2810..4222)
/gene="Rv3450c"
/note="Rv3450c" (MTCY13E12.03c), len: 470. Function:
unknown but similar to Rv3895c MTCY15F10.17 (FASTA score:
1.6e-27; 36.4% identity in 475 aa overlap); possible
membrane spanning region near N-terminus, and to
V14967|MLC628.16 Mycobacterium leprae cosmid H628; (481
aa) opt: 708 Z-score: 710.9 E(): 4.5e-32; 32.9%
identity in 480 aa overlap, also similar to Rv3869
(MTV027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:O06317"
/translation="MPSPATTLVHVSQYRFLLRTECALLFGDVCAATGALRARTSL
ALGCVLATVAMGCAAFVALLRQPSALGQAPIVMGREGALYVRVDVWHPPVNLASAR
LIAATNANPQVSESLGHTKRGPIIGIPAPQLIDQPLAGAESAWAICDSNDCGSTT
VVGPAEDSSAQVLTAEOMILVATFSGSPTVLLYCGRRVAVHAIAPVAVMAIRLQGRV
PHVQAQSLNAVPEAPRIITAPRIHCGGRASVCLPGFLVGVVVKITKASGDEYVYVLED
GVORIGOVAADLLRFGDSQGSVNVPTVADVIRVAPIVNTLVSAFHPKPTVPGKSP
GRAVTLCTVMTFPAQGAARVAFLAGSGPPVPLGGVPTVLAQADGKGALLDAYLPPG
RSAYVAARSLSGGGTGTRYLVTDGVRFAIHDDDDVAHDGLGPTAAIPAPVPLATLPS
GPGLSRANASVARDTVAPCP"
complement(4049..4081)
/gene="Rv3450c"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(4228..4231)
/note="possible RBS, GGAG, for Rv3450c"
4388..5131
/gene="Rv3451"
4388..5131
/gene="Rv3451"
/note="Rv3451" (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MYCTU_Q50664
probable cutinase cy339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1

```

```

/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/db_xref="SWISS-PROT:O06318"
/translation="MGAGALITAVVLLIALGAVMTVAADGCPDAFVTFARCTGSP
GIRGVOAFVDSLRQQTGMHIGVYPVNYAASRLQLHGGDCANAIASHIKSMASCCPNT
KIYLVGYSQCATVNDIVAGVPLGSI1SFGSPLPAAYADNVAAVAFGNPSNRAGCSLSS
LSPLFGSKAIDIAFNPTDPTI"GVGGNEFSGHIDGYIPTTYTQAASEVWOKIRAGSVPH
LPGSVPLQSGSVLQMPCTAPAPESLHGR"
4706..4723
/misc_feature
/gene="Rv3451"
/note="PS00155 Cutinase, serine active site, GYSQG"
5165..5168
/note="possible RBS, GAGG, for Rv3452"
5178..5858
/gene="Rv3452"
5178..5858
/gene="Rv3452"
/note="Rv3452" (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,
Query Match 100.0%; Score 18; DB 1; Length 43401;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 catccgclcatlacty 18
Db 15872 CATGCCGTCGTATTGCTG 15855
|||||
|
RESULT 15
AF176573 9600 bp RNA VRL 18-AUG-1999
LOCUS Hepatitis C virus polyprotein precursor, gene, complete cds.
DEFINITION AF176573
ACCESSION AF176573.1 GI:5738246
VERSION AF176573.1 GI:5738246
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus.
REFERENCE 1 (bases 1 to 9600)
AUTHORS Mokhomov,V.V., Samokhvalov,E.I., Novikov,D.V., Shatalov,A.G. and
Prilipov,A.G.
TITLE Molecular cloning HCV Russian isolate 1b from the serum of patient
with acute hepatitis
JOURNAL unpublished
REFERENCE 2 (bases 1 to 9600)
AUTHORS Mokhomov,V.V., Samokhvalov,E.I., Novikov,D.V., Shatalov,A.G. and
Prilipov,A.G.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1999) Molecular Genetics, Ivanovsky Virology
Institute, Gamaleya Str., 6, Moscow 123098, Russia
FEATURES
location/Qualifiers
1..9600
/organism="Hepatitis C virus"
/strain="274933R0"
/isolate="1b"
/db_xref="taxon:11103"
/country="Russia"
/note="isolated from acute hepatitis patient serum"
1..341
342..9374
/note="cleaved into C, E1, E2, p7, NS2, NS3, NS4a, NS4b,
NS5a, and NS5b proteins"
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAD50312.1"
/db_xref="GI:5738247"
/translation="MSTNPKPQKTKRNTNRPPQDKFPGGQIVGGYVLLPRGPRRL
GVKATRKTSERSOPRQPIPKARHPEGRVWQPGYWPPLYGNPGIGWAGWIIJSPRG
SRPSWGPDPTRRRSRNLKVIDTLTCGFADLMGYIPLVGAPLUGVAKALAHGVRLVED

```

GVNATGNLPCSCSFIFLLALLSCLTIPASAYEVNRNVGYVHVHTNDCSNSSIVYEAD
VIMHTPGVCPCVDQGNTRSCWVALTPTLAARNASVPVIAIRRHVDLLVGTAAFGCSAM
VGLDCGSPVSPVQLETFSPRRHQTVQDCNCSYIPGHISGHRMAMMMNWSPTAALVY
SQLRIQAIYDMVAGAHGVLAYISMGVNAKYMIVILLFAGVDGTHHTTGGAA
ARATQGFSPFSLGQKQIOLINTNGSHNRTALNCNDSLOTGFLAALFYTYRFNAS
GCPERMASCRPIDKFDGOWGPIYAEPPSSDORPCWYAPRPGCIYVPAISOVCGPVYC
ETPSPVYVYVTRDRGCVTYTWGENETDVLNNTRPPLGNWFGCTWNNSTGFTKTCGG
PPCNIGGAGNTLLTCPTDCFRKHPEATYKCGSGPMLTPRCLVDYPIYLMHYPCAVNF
TIFKRYMYGVGHRNLNAACNWTGERKCDLDRDRSELSPLLLSTTEWQVLPSCFTTL
PALSTGLTHQNTVDQYLYIGISAVIPFAIKWEYVLLFLLLADARVACCLMMMLL
IAQAEALLENVLWNAASVAGAHGILSFLVFFCAAMYIKGRVFGAAYALYGVWPLLL
LLALPPRAYAMDRMAAASCGAVFVGLVLLTSPHYKFLARLIIMWSQYFTTRAEAL
LOVWVPLNIIKGRDALLIMCAVHSELIPDITKFLAILILGPLMMFOAIRTRVPIYVR
AQGLLRACNIVRKAAGHYIQMALMKLAALSCTYVYDHLTPILODMAHAGLRDLAAVE
PVFSDMETKITWGAADTAACDIIISGLPVSAARRGRELILGPADGLEEQGWRLLAPIT
AYSQQTRGLLGCIIISLTGRDNQVEGEVQVSTATQSFATCVNGVCWTYVYHAGTK
TLAGPKGPIQMTYNTDQDLVQWQAPPCARSLTPCTCGSSDLYLVTRHADVIYVRRRG
DSRGLSPRPISYLSKSSGGPLPCPSGAVGIFRAAVCTRGVAKAVDFPVESMETT
MRSPVFTDSSPPAVPOTFOVAHLHAPTGGSKTKVPAAYAAQYKVLVLPNSVAATL
GFGAYMSAHGVDPNLTGVRTITTGAPITYSTYKFLADGCGSGGAYDIIICDECHS
TDSITLIGITVLDQATAGARLVVLATATPPGTVTPHPNIEIEIALSNTGEIPEFYK
AIPIEIKGGHRLIFCHSKKCKDELAAGLGLNAYAYHGLDVSVIPTSGNVVVA
DALMTGTFDPSVDCNTCTQTFVSLDPTFTIETTVPODAVSRRQRRTGRG
RGIYREVTPGERSGPMEDSSVLCCEYDAGCAMELT'PAETSVRLRAYLNTPLGLFVQ
DHLREWSVFTGLTHIDAFILSOTKOAGDNFPLVAYOATVCARAQAPPPSWDMWKC
LIRKPTLHGPTLLYRLGAVONETTLT'HPITKYIMACMSADLEVVTSTWLVGGVLA
ALUAYCLTTGSSVIVGRIVLSGKPAIIPDREVLQEFDEMEECASHLPYIBQGMOLAE
QFKKALGLLOATQKAAAPVPSVWRALAEFAWKHMNFISGVQYLAGSLPLGN
PAIASLMAFTASITSPLTQYTLFNILGNGWAAQALAPPASAAGVAGIAGAAGVSI
GLGKLVLDVLAGVAGALVAFKVMSEMPSTEDLVNLLPAIILSPGALYGVVCAA
ILRRHVGEAGAVOMNRLIAFASGNHVSPTHYVPESDAAARVQILSNLTITQLLK
RLHOWNDCSTPCSGSRLDVMWICVLTDFKTLWSKLLPRLPGVPFLSCQGYK
GTVRGDGIWOTPCGCAQIAGHKVNSMRIVGPRTCSNTHGCTP'INAYTTGCTPSP
APNYSRALWRVAEEYVEITRVGDEHYVTGNTDVKPCQVPAPEFTEVDGVRLHR
YAPVCKPDLLEDVTFQVGLNQLVGSQLPCEPEPDVAVLTSMLTDPHSITAEAKRRL
ARGSSPLASSASQLSAPSLKATCTTHDSPDADLIEANULIWMQEMGNGNITRVSEN
KVVIILDSFDPLRAEDEREVSYPAILRKT'KFFSAMP'IMARPDYNPPLLESWKDPDY
VPPVHGGCP'LPPTKVPIPPPRKRKT'VLTETVSSALAEIATKTFGSESAVDSGT
AAASPDQSDNGDTSQDVESEMPPEGEPCDPLSDGSMSTVSEEAEDVCCSMS
YTWGALITPCAAESQLP'INALSNLHRLNLYATT'RSASOROKKVT'FDRLOVLD
DHYQDLKEMKAKASTYKAKLLSVEACKLT'PHSAKSKFCGYGAKDVNLSKAVNHI
HSVKDLEDDETPTIDTTIMAKNEVFCVQPEKGGKPKARLIVFPDGLGVYCEKALYD
VYSTLPQAVMSAYGQYSPQQRVDFLVNAWKKCPMGFSYDTRCFDSTVTESDIRV
EESIYQCCDLAPAEARQAI'RSLT'ERLYVGGPLTNSKGONCYRRCASGLVTTSCNTL
TCYLAKATAACRAAKLRDCTLLVNGDDLVI'ICESAGTQEDAAASLRVFT'EAMT'RYSA
DLPOPEYDLELIT'RSSNVSAHDSAGKRVYVLT'RDPTT'PLARAWEAARHT'PVNSWL
GNIMVAPTLWARMILMT'HFSTILLFQLEKALDCO'LYGAYYSIEPLDLPQIIORLH
GLSAFSLHSYSPEINRVASCLRLK'CVPLRAWHRAHNVAKLILSQGGRAT'CKKYL
FNWAVRTK'KLTP'IPAASQIDLSGMFVAGYGGDIYHSI'SKARPKW'IML'CLILLFPVGV
GIYLLPNR*

3'UTR 9375..9600
BASE COUNT 1917 a 2887 c 2696 g 2100 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 14; Length 9600;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 catgccgtatgctg 18
|||||
Db 1125 CATGACGTCGTATGCTG 1108

Search completed: April 2, 2002, 22:46:14
Job time: 12544 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 08:04:13 ; Search time 165.4 seconds
(without alignments)
24.647 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaaatgattcgcg 18
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	18	100.0	12412	1	US-08-390-878-18
2	16.4	91.1	1535	3	US-08-464-052-1
3	16.4	91.1	1535	4	US-08-461-002-1
4	16.4	91.1	1535	4	US-08-689-411-1
5	16.4	91.1	1535	5	PCT-US94-09863-1
6	16.4	91.1	4411529	4	US-09-103-840A-1
7	14.4	80.0	573	2	US-08-709-912-18
8	14.4	80.0	573	1	US-09-047-370-18
9	14.4	77.8	1959	4	US-09-201-641-1
10	13.8	76.7	358	1	US-08-322-742-19
11	13.8	76.7	2321	4	US-09-198-839-2
12	13.8	76.7	2337	4	US-09-198-839-3
13	13.4	74.4	33	1	US-08-345-756-6
14	13.4	74.4	33	1	US-08-625-198-6
15	13.4	74.4	579	5	PCT-US96-05320A-1664
16	13.4	74.4	859	1	US-08-345-756-8
17	13.4	74.4	859	1	US-08-625-198-8
18	13.4	74.4	1902	3	US-09-041-991A-7
19	13.4	74.4	2039	1	US-08-345-756-5
20	13.4	74.4	2039	1	US-08-625-198-5
21	13.4	74.4	5467	2	US-08-605-106-7
22	13.2	73.3	30	1	US-08-547-182-4
23	13.2	73.3	60	3	US-08-577-121-3
24	13.2	73.3	60	4	US-08-985-700-3
25	13.2	73.3	60	4	US-09-149-161-3
26	13.2	73.3	60	5	PCT-US95-16916-3
27	13.2	73.3	1001	3	US-08-096-044C-2

28	13.2	73.3	1112	3	US-08-096-044C-9	Sequence 9, Appli
29	13.2	73.3	1530	1	US-08-149-105-1	Sequence 1, Appli
30	13.2	73.3	1530	1	US-08-317-847-1	Sequence 1, Appli
31	13.2	73.3	1557	5	PCT-US91-01327-11	Sequence 11, Appli
32	13.2	73.3	1557	6	5248670-1	Patent No. 5248670
33	13.2	73.3	1882	1	US-08-431-080-15	Sequence 15, Appli
34	13.2	73.3	1882	1	US-08-441-080-29	Sequence 29, Appli
35	13.2	73.3	1882	2	US-08-938-534-15	Sequence 15, Appli
36	13.2	73.3	1882	2	US-08-938-534-29	Sequence 29, Appli
37	13.2	73.3	4238	4	US-08-124-934A-5	Sequence 5, Appli
38	13.2	73.3	4238	5	PCT-US94-10080-5	Sequence 5, Appli
39	13.2	73.3	5211	1	US-08-447-411-1	Sequence 1, Appli
40	13.2	73.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
41	13.2	73.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
42	13.2	73.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
43	13	72.2	39	3	US-08-641-873-17	Sequence 17, Appli
44	13	72.2	1742	4	US-09-099-676-2	Sequence 2, Appli
45	13	72.2	1742	4	US-09-565-910-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-390-878-18
; Sequence 18, Application US/08390878
; Patent No. 5700684
; GENERAL INFORMATION:
; APPLICANT: Stovel, Charles K.
; APPLICANT: Mahalinas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Townsend Klorie and Crew
; STREET: One Market Plaza, Stewart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390.878
; FILING DATE: 17 FEB 1995
; CLASSIFICATION: 15
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Jan
; REGISTRATION NUMBER: 48,498
; REFERENCE/BOOKET NUMBER: 15471A 1/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/54459600
; TELEFAX: 415/5445043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 100.0%; Score 18; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acatcaaatgattcgcg 18
DB 468 ACATCAAAAGTGATTCGG 485

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-1

Query Match 91.1%; Score 16.4; DB 4; Length 1535;
Best Local Similarity 94.4%; Pred. No. 2.6;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgcg 18
|| |||||
DB 841 ACGTCAAGTGATTCGCG 858

RESULT 4
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-689-411-1

RESULT 2
US-08-464-052-1
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-464-052-1

Query Match 91.1%; Score 16.4; DB 3; Length 1535;
Best Local Similarity 94.4%; Pred. No. 2.6;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgcg 18
|| |||||
DB 841 ACGTCAAGTGATTCGCG 858

RESULT 3
US-08-461-002-1
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgcgc 17
Db 471 CATTAAGTGATTCGC 456

RESULT 8

US-09-047-370-18/c
; Sequence 18, Application US/09047370

; Patent No. 5866408

; GENERAL INFORMATION:

; APPLICANT: Sung Dr., Wing L

; APPLICANT: Yaguchi Dr., Makoto

; APPLICANT: Ishikawa Dr., Kazuhiko

; TITLE OF INVENTION: Modification of Xylanase to Improve

; TITLE OF INVENTION: Thermophilicity, Alkalophilicity and

; TITLE OF INVENTION: Thermostability

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto

; STREET: 277 Park Ave.

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10172-0194

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: -US/09/047,370

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/709,912

; FILING DATE: 09-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Olsen Mr, Warren E

; REGISTRATION NUMBER: 27290

; REFERENCE/DOCKET NUMBER: 1039,2000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-2400

; TELEFAX: (212) 758-2982

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 573 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Synthetic DNA"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; IMMEDIATE SOURCE:

; CLONE: pTVX(3-190)

US-09-047-370-18

Query Match 80.0%; Score 14.4; DB 2; Length 573;

Best Local Similarity 93.8%; Pred. NO. 30;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgcgc 17

Db 471 CATTAAGTGATTCGC 456

RESULT 9

US-09-201-641-1

; Sequence 1, Application US/09201641A

; Patent No. 6232530

;

; GENERAL INFORMATION:

; APPLICANT: Cunningham Jr, Francis X

; APPLICANT: DellaPenna, Dean

; TITLE OF INVENTION: Method for Regulating Carotenoid Biosynthesis in

; TITLE OF INVENTION: Marigolds

; FILE REFERENCE: Quest 41-162

; CURRENT APPLICATION NUMBER: US/09/201,641A

; CURRENT FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1959

; TYPE: DNA

; ORGANISM: Tagetes erecta

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (304)...(1836)

; OTHER INFORMATION: beta-cyclase

US-09-201-641-1

Query Match 77.8%; Score 14; DB 4; Length 1959;

Best Local Similarity 100.0%; Pred. NO. 61;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtgatgc 15

Db 897 catcaaaagtgatgc 910

RESULT 10

US-08-322-742-19

; Sequence 19, Application US/08322742

; Patent No. 5688641

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 50Z or 55SX

; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/322,742

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/948,823

; FILING DATE: September 1, 1992

; APPLICATION NUMBER: 07/844,296

; FILING DATE: February 28, 1992

; APPLICATION NUMBER: 07/552,216

; FILING DATE: February 28, 1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 00530/048003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 358

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear
US-08-322-742-19

Query Match 76.7%; Score 13.8; DB 1; Length 358;
Best Local Similarity 88.2%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
||||| ||||||| ||
Db 262 CATCACAGTGATTCGCG 278

RESULT 11
US-09-198-839-2
; Sequence 2, Application US/09198839
; Patent No. 6190908
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray S.
; TITLE OF INVENTION: MODULATION OF POLYPEPTIDE DISPLAY ON MODIFIED
; FILE OF INVENTION: FILAMENTOUS PHASE
; FILE REFERENCE: SCR21115
; CURRENT APPLICATION NUMBER: US/09/198,839
; CURRENT FILING DATE: 1998-12-24
; PRIOR APPLICATION NUMBER: PCT/US99/18207
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 60/096,326
; PRIOR FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2321
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PORFES II
US-09-198-839-2

Query Match 76.7%; Score 13.8; DB 4; Length 2321;
Best Local Similarity 88.2%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
||||| ||||||| ||
Db 2043 catcaaatgttcgcg 2059

RESULT 12
US-09-198-839-3
; Sequence 3, Application US/09198839
; Patent No. 6190908
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray S.
; TITLE OF INVENTION: MODULATION OF POLYPEPTIDE DISPLAY ON MODIFIED
; FILE OF INVENTION: FILAMENTOUS PHASE
; FILE REFERENCE: SCR21115
; CURRENT APPLICATION NUMBER: US/09/198,839
; CURRENT FILING DATE: 1998-12-24
; PRIOR APPLICATION NUMBER: PCT/US99/18207
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 60/096,326
; PRIOR FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2337
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PORFES IV
US-09-198-839-3

Query Match 76.7%; Score 13.8; DB 4; Length 2337;
Best Local Similarity 88.2%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
||||| ||||||| ||
Db 2059 catcaaatgttcgcg 2075

RESULT 13
US-08-345-756-6
; Sequence 6, Application US/08345756
; Patent No. 5633438
; GENERAL INFORMATION:
; APPLICANT: Hasczynski, Chris
; APPLICANT: Harbour, Eric
; APPLICANT: Hattori, Jiro
; APPLICANT: Miki, Brian
; TITLE OF INVENTION: MICRUSPORE-SPECIFIC REGULATORY ELEMENT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,756
; FILING DATE: 22-NOV-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, STEPHEN A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/236/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (circular)
US-08-345-756 6

Query Match 74.4%; Score 13.4; DB 1; Length 33;
Best Local Similarity 93.3%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 16
||||| ||||||| ||
Db 12 CATCAAACTGATGCG 26

RESULT 14
US-08-625-198-6
; Sequence 6, Application US/08625198
; Patent No. 5756324
; GENERAL INFORMATION:
; APPLICANT: Hasczynski, Chris
; APPLICANT: Harbour, Eric
; APPLICANT: Hattori, Jiro
; APPLICANT: Miki, Brian

;; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/625,198
;; FILING DATE: 01-APR-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/345,756
;; FILING DATE: 22-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, STEPHEN A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 33229/236/PIH1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 33 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-625-198-6

Query Match 74.4%; Score 13.4; DB 1; Length 33;
Best Local Similarity 93.3%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtatttcg 16
|||||
Db 12 CATCAAAAGTGATCG 26

RESULT 15
PCT-US96-05320A-1664
; Sequence 1664, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US96/05320A
;; FILING DATE: April 22, 1996
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/476,102
;; FILING DATE: June 7, 1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/487,429
;; FILING DATE: June 7, 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Eric K. Stoffe
;; REGISTRATION NUMBER: 36,688
;; REFERENCE/DOCKET NUMBER: 1488.014PC01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 1664:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 579 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: Linear
PCT-US96-05320A-1664

Query Match 74.4%; Score 13.4; DB 5; Length 579;
Best Local Similarity 93.3%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tcacaaatgatttcg 18
|||||
Db 446 TCACAAAGTGATGCG 460

Search completed: April 3, 2002, 08:07:11
Job time: 42191 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 21:47:57 ; Search time 3835.7 Seconds
(without alignments)
50.427 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaagtgattcgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estirr:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	16	88.9	135	10	AA249381	AA249381 j3441.seq
2	16	88.9	160	11	BF351466	BF351466 OVI-HT051
3	16	88.9	356	11	BF376314	BF376314 CM0-TN003
4	16	88.9	375	11	BF367020	BF367020 PM0-GN001
5	16	88.9	407	11	BF367003	BF367003 PM0-GN001
6	16	88.9	415	11	D44620	D44620 HUMSUPY017
7	16	88.9	427	11	BF772419	BF772419 IL5-IT002
8	16	88.9	594	10	AW750028	AW750028 PM2-BT054
9	15.4	85.6	166	11	BF944885	BF944885 CM1-NN019
10	15.4	85.6	238	10	AW248379	AW248379 2820521.5
11	15.4	85.6	269	11	F05463	F05463 HSCOBH041 n
12	15.4	85.6	301	11	F12193	F12193 HSC36H081 n

13	15.4	85.6	345	11	H14295	H14295 ym63b06.r1
14	15.4	85.6	360	11	BF947972	BF947972 CM1-NN019
15	15.4	85.6	380	11	EG639721	EG639721 SD04846.3
16	15.4	85.6	382	11	EG639722	EG639722 SD04865.3
17	15.4	85.6	383	11	BF947969	BF947969 CM1-NN019
18	15.4	85.6	385	10	AL134134	AL134134 DKF2P547J
19	15.4	85.6	414	11	BF944889	BF944889 CM1-NN019
20	15.4	85.6	415	11	BF927805	BF927805 IL5-NTU22
21	15.4	85.6	416	11	BF944887	BF944887 CM1-NN019
22	15.4	85.6	467	11	BF490411	BF490411 AT26869.5
23	15.4	85.6	474	13	AQ692029	AQ692029 HS_5408_A
24	15.4	85.6	499	10	AA141880	AA141880 CK02646.3
25	15.4	85.6	508	10	AW390923	AW390923 MR3-STU20
26	15.4	85.6	524	13	AQ928124	AQ928124 KPC1-24-2
27	15.4	85.6	585	10	AL543240	AL543240 AL543240
28	15.4	85.6	606	11	BF238039	BF238039 601811750
29	15.4	85.6	610	10	AF513102	AF513102 GH01393.3
30	15.4	85.6	631	11	BF718674	BF718674 602696791
31	15.4	85.6	656	11	BF347028	BF347028 602021824
32	15.4	85.6	854	10	AL520879	AL520879 AL520879
33	15.4	85.6	885	10	AL559329	AL559329 AL559329
34	15.4	85.6	886	10	BF735986	BF735986 601405458
35	15.4	85.6	897	10	AL556646	AL556646 AL556646
36	15.4	85.6	916	11	BI115704	BI115704 602866051
37	15.4	85.6	940	10	AL536089	AL536089 AL536089
38	15.4	85.6	962	13	CNS02RLU	AL210747 Tetraodon
39	15.4	85.6	968	10	AL525534	AL525534 AL525534
40	15.4	85.6	972	10	AL551791	AL551791 AL551791
41	15.4	85.6	1094	11	W18183	W18183 IMAGE:20087
42	15	83.3	227	10	BE240803	BE240803 EST404852
43	15	83.3	237	11	EG382488	EG382488 298399 MA
44	15	83.3	267	10	AW774157	AW774157 EST333240
45	15	83.3	295	13	AQ091462	AQ091462 HS_3016_B

ALIGNMENTS

RESULT 1

AA249381

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

j3441.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
5', mRNA sequence.
AA249381.1 GI:1880285
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Liew,C.C.
CDNAs from human fetal heart (1997)
Unpublished (1997)
Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
PCR Primers
FORWARD: 5' GCCAAGCTCGAATTACCCCTCACTAAAGGG 3'
BACKWARD: 5' CCACGTGATTCGTAATAGCTACTATAGGCG 3'
Seq primer: 5' GAATTAACCCCTCACTAAAGGG 3'.
Location/Qualifiers
1. .135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site1: EcoRI; Site2:
XhoI; mRNA was purified from human fetal hearts (8-10

FEATURES

source

weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 36 a 43 c 33 g 23 t
ORIGIN

Query Match 88.9%; Score 16; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
|||||

Db 28 CATCAAGTGATTCGC 43

RESULT 2
BF351466 160 bp mRNA EST 22-NOV-2000
LOCUS QV1-HT0517-020400-142-a08 HT0517 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF351466
ACCESSION BF351466
VERSION BF351466.1 GI:11310540
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 160)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV1-HT0517-020400-142-a08&t3=2000-04-02&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers
1..160
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0517"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 46 a 48 c 40 g 26 t
ORIGIN

Query Match 88.9%; Score 16; DB 11; Length 160;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
|||||

Db 20 CATCAAGTGATTCGC 35

RESULT 3

BF376314 356 bp mRNA EST 24-NOV-2000
LOCUS CM0-TN0039-210800-507-c05 TN0039 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF376314
ACCESSION BF376314.1 GI:11338339
VERSION EST.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Mtheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 356)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-CM0&t2-CM0-TN0039-210800-507-c05&t3=2000-08-21&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 59
High quality sequence stop: 356.
Location/Qualifiers
1..356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TN0039"
/dev_stage="Adult"
/note="Organ: testis normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES
source

BASE COUNT 89 a 95 c 106 g 66 t
ORIGIN

Query Match 88.9%; Score 16; DB 11; Length 356;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
|||||

Db 74 CATCAAGTGATTCGC 89

RESULT 4

BF367020 375 bp mRNA EST 24-NOV-2000
LOCUS PM0-GNU018-130900-003-e09 GNU018 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF367020
ACCESSION BF367020


```

VERSION      BF367020.1  GI:11329071
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 375)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-GN0018-
             130900-003-e09&t3=2000-09-13&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 375.
             Location/Qualifiers
             1..375-
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="GN0018"
             /dev_stage="Adult"
             /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
             ; Site_2: SmaI; A mini-library was made by cloning
             products derived from ORESTES PCR (U.S. Letters Patent
             application No. 196,716 - Ludwig Institute for Cancer
             Research) profiles into the puc 18 vector. Reverse
             transcription of tissue mRNA and cDNA amplification were
             performed under low stringency conditions."
             87 a 98 c 106 g 84 t

BASE COUNT   87 a 98 c 106 g 84 t
ORIGIN
Query Match      88.9%; Score 16; DB 11; Length 375,
Best Local Similarity 100.0%; Pred. NO. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaagtgattgcg 17
|||||
Db 311 CATCAAGTGATTGCG 326

RESULT 5
LOCUS      BF367003 407 bp mRNA EST 24-NOV-2000
DEFINITION PM0-GN0018-040900-002-A08 GN0018 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF367003
VERSION     BF367003.1 GI:11329054
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 407)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-GN0018-
             130900-003-e09&t3=2000-09-13&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 375.
             Location/Qualifiers
             1..407
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="GN0018"
             /dev_stage="Adult"
             /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
             ; Site_2: SmaI; A mini-library was made by cloning
             products derived from ORESTES PCR (U.S. Letters Patent
             application No. 196,716 - Ludwig Institute for Cancer
             Research) profiles into the puc 18 vector. Reverse
             transcription of tissue mRNA and cDNA amplification were
             performed under low stringency conditions."
             95 a 108 c 110 g 93 t 1 others

BASE COUNT   95 a 108 c 110 g 93 t 1 others
ORIGIN
Query Match      88.9%; Score 16; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. NO. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaagtgattgcg 17
|||||
Db 329 CATCAAGTGATTGCG 344

RESULT 6
LOCUS      D44620 415 bp mRNA EST 20-FEB-1998
DEFINITION HUMSHIP017 human brain cDNA Homo sapiens cDNA clone 003, mRNA
sequence.
ACCESSION  D44620
VERSION     D44620.1 GI:1572095
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 415)
AUTHORS      Hadano,S., Ishida,Y., Tomiyasu,H., Yamamoto,K., Bates,G.P. and
             Ikeda,J.
TITLE        Transcript map of the human chromosome 4p16.3 consisting of 627
             cDNA clones derived from 1 Mb of the Huntington's disease locus
             DNA Res. 3 (4), 239-255 (1996)
JOURNAL      97101646
MEDLINE      97101646
COMMENT      Contact: Shinji Hadano
             Japan Science and Technology Corporation, NeuroGenes Project, ICORP
             Univ. of Tokai School of Med.
             Boiseidai, Isehara, Kanagawa 259-1193, Japan
             Tel: 81-463-91-5095
             Fax: 81-463-91-4993
             Email: shinji@ngs.med.u-tokai.ac.jp

```

```

FEATURES
Source
    Location/Qualifiers
    1. 415
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="003"
    /clone_lib="Human brain cDNA"
    /tissue_type="brain"
BASE COUNT      81 a 109 c 110 g 114 t 1 others
ORIGIN

Query Match      88.9%; Score 16; DB 11; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattcgc 17
    |||||
Db 360 CATCAAAAGTGATTCGC 375

RESULT 7
BF772419
LOCUS      BF772419      427 bp      mRNA      12-JAN-2001
DEFINITION IL5-IT0027-121200-322-b06 IT0027 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF772419
VERSION    BF772419.1 GI:12120319
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 427)
AUTHORS  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT  Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=115&t2=115-IT0027-
    121200-322-b06&t3=2000-12-12&t4=1)
    Seq primer: puc 18 forward
    High quality sequence stop: 400.
    Location/Qualifiers
    1. 427
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="IT0027"
    /dev_stage="Adult"
    /note="Organ: epid_tumor; Vector: puc18; Site_1: Smal;
    Site_2: Smal; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT      120 a 113 c 120 g 71 t 3 others
ORIGIN

FEATURES
Source
    Location/Qualifiers
    1. 415
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="Human brain cDNA"
    /tissue_type="brain"
BASE COUNT      81 a 109 c 110 g 114 t 1 others
ORIGIN

Query Match      88.9%; Score 16; DB 11; Length 427;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattcgc 17
    |||||
Db 432 CATCAAAAGTGATTCGC 417

RESULT 8
AW750028/c
LOCUS      AW750028      594 bp      mRNA      EST      28-APR-2000
DEFINITION PM2-BT0546-281299-003-107 BT0546 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW750028
VERSION    AW750028.1 GI:7664960
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 594)
AUTHORS  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT  Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-BT0546-
    281299-003-107&t3=1999-12-28&t4=1)
    Seq primer: puc 18 forward
    High quality sequence stop: 581.
    Location/Qualifiers
    1. 594
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0546"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: Smal; Site_2:
    Smal; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the pUC 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT      109 a 169 c 160 g 156 t
ORIGIN

Query Match      88.9%; Score 16; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattcgc 17
    |||||
Db 432 CATCAAAAGTGATTCGC 417

RESULT 9

```

BF944885 166 bp mRNA EST 22-JAN-2001
 CM1-NN0193-191000-484-g08 NN0193 Homo sapiens cDNA, mRNA sequence.
 BF944885
 BF944885.1 GI:12362160
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 166)
 Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva W. Jr., Zago, M.A., Bordin, S., Costa, F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&t2=CM1-NN0193-191000-484-g08&t3=2000-10-19&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 16
 High quality sequence stop: 166.
 Location/Qualifiers
 1. .166
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0193"
 /envi_stage="Adult"
 /note="Organ: nervous_normal; Vector: puc18; Site_1: Smal;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 41 a 44 c 50 g 31 t
 BASE COUNT
 ORIGIN
 Query Match 85.6%; Score 15.4; DB 11; Length 166;
 Best Local Similarity 94.1%; Pred. No. 7.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 acatcaaaagtgttcgc 17
 |||||
 Db 81 ACATCAAAAGTGATTGCC 97
 RESULT 10
 AW248379
 LOCUS
 DEFINITION
 AW248379
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-0bh04
Seq primer: (-21)M13 universal.

FEATURES

source Location/Qualifiers

1..269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-0bh04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 65 a 75 c 90 g 39 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 269;
Best Local Similarity 94.1%; Pred. No. 7.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtattcgc 17

|||||

Db 137 ACATCAAAAGTGATGGC 153

RESULT 12

F12193 301 bp mRNA EST 13-MAR-1995
LOCUS HSC36H081 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION C-36H08, mRNA sequence.

ACCESSION F12193

VERSION F12193.1 GI:706535

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houllgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouillot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.

IMAGE: molecular integration of the analysis of the human genome

and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL 95277534

MEDLINE

COMMENT

Contact: Genethon

Genexpress-Genethon

Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpress@genethon.fr

Single read.

Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-36h08

Seq primer: (-21)M13 universal.

Location/Qualifiers

1..301

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="c-36h08"

/clone_lib="normalized infant brain cDNA"

/sex="Female"

/tissue_type="total brain"

/dev_stage="3 months old"

/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 73 a 78 c 99 g 47 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 301;
Best Local Similarity 94.1%; Pred. No. 7.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtattcgc 17

|||||

Db 135 ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

|||||

ACATCAAAAGTGATGGC 151

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAACTGAGTGGGAGCGCGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M.Fatima Bonaldo. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT 90 a 81 c 100 g 71 t 3 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 345;
Best Local Similarity 94.1%; Pred. No. 8.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgttcgc 17

Db 73 ACATCAAAAGTGTTCGC 89

RESULT 14

BF947972/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

LABORATORY

ADDRESS

TELEPHONE

FAX

EMAIL

PROJECT

URL

SEQUENCE

PRIMER

High quality sequence start: 35

High quality sequence stop: 360.

Location/Qualifiers

1. .360

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NN0193"

/dev_stage="Adult"

/note="organ: nervous_normal; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 79 a 99 c 91 g 91 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 360;
Best Local Similarity 94.1%; Pred. No. 8.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgttcgc 17

Db 334 ACATCAAAAGTGTTCGC 318

RESULT 15

BG639721/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

LABORATORY

ADDRESS

TELEPHONE

FAX

EMAIL

PROJECT

URL

SEQUENCE

PRIMER

High quality sequence start: 209.

Location/Qualifiers

1. .380

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="SD04846"

/clone_lib="SD04846"

/lab_host="DH5-alpha"

/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized

fractionated cDNAs were directly ligated into pOT2.

Plasmid cDNA library."

Db 379 CATCAAGAGAGATTCGGC 363
||||||| |||||||

Search completed: April 2, 2002, 21:48:01
Job time: 10826 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:49:16 ; Search time 366.19 Seconds
(without alignments)
42.142 Million cell updates/sec

Title: US-09-785-904-4
Perfect score: 18
Sequence: 1 catgccgtcgtattgctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

- 1: /SIDS2/gcgdata/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	100.0	650	17 AAT33658	M. tuberculosis ma
C 2	18	100.0	650	19 AAV18649	DNA for M. tubercu
C 3	18	100.0	1535	16 AAQ89200	Mycobacterium tube
C 4	18	100.0	1535	17 AAT33656	M. tuberculosis ce
C 5	18	100.0	1535	19 AAV18647	DNA for M. tubercu
C 6	18	100.0	12412	17 AAT33537	BCG deletion regio
7	15.4	85.6	1368	21 AAZ53670	Neisseria gonorrhoe
8	15.4	85.6	1779	20 AAZ12075	Neisseria gonorrhoe
9	15.4	85.6	1779	21 AAZ53673	Neisseria gonorrhoe
10	15.4	85.6	5859	17 AAT14061	N. gonorrhoeae lgt
11	15.4	85.6	5859	18 AAT49230	Lipo-oligosacchari

C 12	15.4	85.6	27425	21	AAA53978	DMP53 tumour suppr
C 13	15	83.3	742	13	AAO20926	C10-E15 DNA fragme
C 14	15	83.3	932	13	AAO20923	C10-E12 DNA fragme
15	15	83.3	1569	20	AAI19484	Streptococcus pneu
16	15	83.3	1572	22	AAH90724	CFE 28 coding sequ
17	15	83.3	1605	22	AAH90837	2CFE 28 coding seq
C 18	15	83.3	2026	19	AAH96317	S. pneumoniae deri
C 19	15	83.3	2026	19	AAH96317	Streptococcus pneu
C 20	15	83.3	7577	19	AAV52179	Human colon cancer
C 21	14.8	82.2	478	22	AAH34918	Hepatitis C virus
C 22	14.8	82.2	576	16	AAQ83862	Hepatitis C virus
C 23	14.8	82.2	576	17	AAT16576	Arabidopsis thalia
C 24	14.8	82.2	1317	21	AAI34516	Mouse neuropeptide
C 25	14.8	82.2	1585	18	AAI766912	DNA encoding human
C 26	14.8	82.2	3086	22	AAE27649	Non-A, non-B hepat
C 27	14.8	82.2	3401	15	AAQ64069	Human cytokine alp
C 28	14.8	82.2	3671	22	AAT30387	Human cytokine alp
C 29	14.8	82.2	105973	21	AAE22298	BAC containing reg
C 30	14.8	82.2	577	14	AAQ35085	HCV envelope prote
C 31	14.4	80.0	780	21	AAI12350	Aspergillus oryzae
C 32	14.4	80.0	840	20	AAI51725	DNA encoding a hum
C 33	14.4	80.0	983	12	AAQ13846	Phibacin PSBX orfl
C 34	14.4	80.0	1200	12	AAQ13848	Phibacin PSBX orfl
C 35	14.4	80.0	4615	15	AAQ70724	TATA-binding prote
C 36	14.4	80.0	4615	15	AAQ70724	TATA-binding prote
C 37	14.4	80.0	4615	17	AAI42210	Drosophila TATA-bi
C 38	14.4	80.0	4615	18	AAT79604	Zea mays DNA fragm
C 39	14	77.8	1451	21	AAC44048	Neisseria gonorrhoe
C 40	14	77.8	2178	21	AAZ53939	MEK33 cDNA Mus S
C 41	14	77.8	3089	16	AAI05572	Probe #13262 for g
C 42	13.8	76.7	189	22	AAI23329	Probe #17336 used
C 43	13.8	76.7	189	22	AAI48650	Murine bone morpho
C 44	13.8	76.7	272	16	AAQ96221	Murine homologue t
C 45	13.8	76.7	272	16	AAQ96220	

ALIGNMENTS

RESULT 1

AAT33658/C

ID AAT33658 standard; DNA; 650 BP.

XX

AC AAT33658;

XX

DT 22-NOV-1996 (first entry)

XX

DE M tuberculosis macrophage survival gene.

XX

KW Cellular uptake; cell entry; macrophage survival; vaccine;

XX

KW Passive immunisation; gene therapy; ds.

XX

OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).

XX

PN WO9626275-A1.

XX

PD 29-AUG-1996.

XX

PF 20-FEB-1996; 96WO-US02155.

XX

PR 22-FEB-1995; 95US-0392210.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Riley LW;

XX

DR WPI; 1996-425086/42.

XX

PT P-PSDB; AAQ02303.

XX

PT DNA giving M. tuberculosis ability to enter mammalian cells

XX

PT and/or survive within macrophage(s), useful in vaccines to protect

XX

PT mammals against Mycobacterium tuberculosis infection

PS Claim 7; Page 45; 67pp; English.

XX A DNA molecule (AA02303) codes for a protein (AA02303) that confers
CC on Mycobacterium tuberculosis an ability to survive within
CC macrophages. It is a fragment of a larger molecule (AA02303) obtd.
CC by ligating M. tuberculosis genomic DNA fragments into pBluescript
CC II vector and screening recombinant E. coli strains for Hela cell-
CC invasive clones. It can be incorporated into a vector and used
CC for prodn. of recombinant macrophage survival protein, which is
CC useful in vaccines or for facilitating uptake of other materials,
CC e.g. therapeutic genes, into mammalian cells. A second DNA molecule
CC (AA02303) codes for a protein (AA02303) that confers on M.
CC tuberculosis the ability to enter mammalian cells.

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 650;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtattgctg 18
|||||

Db 189 CATGCCGTCGTATTGCTG 172

RESULT 2

AAV18649/c
ID AAV18649 standard; DNA; 650 BP.

XX AAV18649;

XX 03-JUL-1998 (first entry)

XX DNA for M. tuberculosis cellular uptake protein fragment.

XX Cellular uptake protein; vaccine; infection; ds.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 1..649

FT CDS /*tag= a

FT /*note= "stop codon not given"

XX WO9805784-A1.

XX 12-FEB-1998.

XX 06-AUG-1997; 97WO-US13056.

XX 07-AUG-1996; 96US-0689411.

XX (CONN-) CONNAUGHT LAB LTD.
PA (CORR) CORNELL RES FOUND INC.

XX Chong P, Riley LW;

XX WPI; 1998-145620/13.

XX P-PSDB; AAW47543.

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
XX therapeutic agents e.g. antibiotics, also useful in vaccines

XX Disclosure; Page 14; 82pp; English.

XX The present sequence encodes a Mycobacterium tuberculosis cellular
XX uptake protein fragment, which confers on M. tuberculosis an
XX ability to survive within macrophages.

XX The protein can be used in a vaccine to prevent M. tuberculosis
XX infection, and provide for the uptake in cells of, e.g.

XX antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
XX raised against it can be used to treat mammals already exposed to

CC M. tuberculosis, to induce a passive immunity and prevent disease
CC occurrence.

XX Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 650;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtattgctg 18

|||||

Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3

AAQ89200/c

ID AAQ89200 standard; DNA; 1535 BP.

XX AAQ89200;

DT 09-OCT-1995 (first entry)

XX Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
DE protein.

XX Vaccine; tuberculosis; ss.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 1..1535

FT CDS /*tag= a

XX WO9506726-A.

XX 09-MAR-1995.

XX 01-SEP-1994; 94WO-US09863.

XX 02-SEP-1993; 93US-0118442.

XX (CORR) CORNELL RES FOUND INC.

XX Riley LW;

XX WPI; 1995-115442/15.

XX P-PSDB; AAR71931.

XX DNA encoding for cellular uptake of Mycobacterium tuberculosis -
XX used to develop prods for vaccines, passive immunisation and
XX diagnosis and cellular uptake of other materials

XX Claim 2; Page 9-11; 46pp; English.

XX The isolated DNA molecule of the invention confers on M.

XX tuberculosis an ability to enter cells and to survive within

XX macrophages. It encodes a polypeptide having a mol. wt. of about 50-
XX 55 kDa, pref. 52 kDa. The AA sequence represents a highly

XX hydrophilic protein with a hydrophobic region at its carboxy

XX terminus. It could be a secreted protein, a cytoplasmic protein, or

XX a surface protein with its carboxy terminus attached to the outer
XX membrane of the organism. The deduced AA sequence is in AAR71931.

XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 100.0%; Score 18; DB 16; Length 1535;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtattgctg 18

|||||

Wed Apr 3 08:33:27 2002

Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 4

AAAT33656/c

ID AAT33656 standard; DNA; 1535 BP.

XX AC AAT33656;

XX DT 22-NOV-1996 (first entry)

XX DE M. tuberculosis cellular uptake gene region.

XX KW Cellular uptake; cell entry; macrophage; passive immunisation;

XX KW vaccine; gene therapy; ds.

XX OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).

XX FH Key Location/Qualifiers

XX CDS 181..810

XX FT /*tag= a

XX FT /label= ORF-1

XX FT /product= cell entry protein

XX FT 886..1535

XX FT /*tag= b

XX FT /label= ORF-2

XX FT /product= macrophage survival protein

XX PN W09626275-A1.

XX PD 29-AUG-1996.

XX PF 20-FEB-1996; 96WO-US02155.

XX PR 22-FEB-1995; 95US-0392210.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Riley LW;

XX DR WPI; 1996-425086/42.

XX DR P-PSDB; AAW02301.

XX PT DNA giving M. tuberculosis ability to enter mammalian cells -

XX PT and/or survive within macrophage(s), useful in vaccines to protect

XX PT mammals against Mycobacterium tuberculosis infection

XX PS Claim 2; Page 41; 67pp; English.

XX CC A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an

XX CC ability to enter mammalian cells and to survive within macrophages.

XX CC The encoded protein sequence is given in AAW02301. The DNA was obtd.

XX CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II

XX CC vector and screening recombinant E. coli strains for beta cell-

XX CC invasive clones. The DNA includes 2 separate coding regions (see

XX CC also AAT33657-58) coding for the cell entry (AAW02302) and macrophage

XX CC survival (AAW02303) proteins. It can be used to produce the cellular

XX CC uptake proteins used as vaccines or to facilitate uptake of other

XX CC materials, e.g. therapeutic genes.

XX CC Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

XX CC Query Match 100.0%; Score 18; DB 17; Length 1535;

XX CC Best Local Similarity 100.0%; Pred. No. 2;

XX CC Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 catgccgtcgattgctg 18

XX Db 1074 CATGCCGTCGTATTGCTG 1057

XX RESULT 5

AAV18647/c

ID AAV18647 standard; DNA; 1535 BP.

XX AC

XX AAV18647;

XX DT 03-JUL-1998 (first entry)

XX DE DNA for M. tuberculosis cellular uptake protein fragment.

XX KW Cellular uptake protein; vaccine; infection; ds.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers

XX CDS 1..1534

XX FT /*tag= a

XX FT /note= "stop codon not given"

XX PN W09805784-A1.

XX PD 12-FEB-1998.

XX PF 06-AUG-1997; 97WO-US13056.

XX PR 07-AUG-1996; 96US-0689411.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Chong P, Riley LW;

XX DR WPI; 1998-145620/13.

XX DR P-PSDB; AAW47541.

XX PT Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
therapeutic agents e.g. antibiotics, also useful in vaccines

XX PS Disclosure; Pages 9-10; 82pp; English.

XX CC The present sequence encodes a Mycobacterium tuberculosis cellular
uptake protein fragment, which confers on M. tuberculosis an
ability to enter mammalian cells and to survive within macrophages.XX CC The protein can be used in a vaccine to prevent M. tuberculosis
infection, and provide for the uptake in cells of, e.g.
antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
raised against it can be used to treat mammals already exposed to
M. tuberculosis, to induce a passive immunity and prevent disease
occurrence.

XX CC Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

XX CC Query Match 100.0%; Score 18; DB 19; Length 1535;

XX CC Best Local Similarity 100.0%; Pred. No. 2;

XX CC Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 catgccgtcgattgctg 18

XX Db 1074 CATGCCGTCGTATTGCTG 1057

XX RESULT 6

AAAT33537/c

ID AAT33537 standard; DNA; 12412 BP.

XX AC AAT33537;

XX DT 15-FEB-1998 (first entry)

XX DE BCG deletion region 3 and flanking sequences.

XX KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;

XX KW mycobacteria; vaccine; infection; marker; ss.

```

XX OS Mycobacterium bovis strain BCG.
XX FH Key Location/Qualifiers
XX FT 1406..10673
XX FT misc_feature /*tag= a
XX FT /*note= "BCG delta 1 deletion region"
XX PN WO9625519-A1.
XX PD 22-AUG-1996.
XX PF 15-FEB-1996; 96WO-US01938.
XX PR 17-FEB-1995; 95US-0390878.
XX PA (PATH-) PATHOGENESIS CORP.
XX PI Mahairas GG, Stover CK;
XX WPI; 1996-393419/39.
XX Detecting markers for avirulence in Mycobacterium - used in
PT production of vaccines against bacterial infection, and to detect
PT bacterial infection
XX Example 1; Fig 3; 66pp; English.
XX This DNA sequence comprises Mycobacterium bovis BCG deletion
CC sequence BCGdelta3. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 Other
CC deletion regions (see AAT33535 and AAT33536) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absence of deletion junctions (see AAT33338-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
XX SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;

Query Match 100.0%; Score 18; DB 17; Length 12412;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtcgtattgctg 18
Db 705 CATGCCGTCGTATGCTG 688

RESULT 7
AAZ53670
ID AAZ53670 standard; DNA; 1368 BP.
XX AC AAZ53670;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 311 partial DNA sequence SEQ ID NO:1289.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO9957280-A2.

```

```

XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarcellini M;
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR P-PSDB; AAY74908.
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX Claim 7; Page 709; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present inventions. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX SQ Sequence 1368 BP; 287 A; 377 C; 448 G; 256 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 1468;
Best Local Similarity 94.1%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgcgcgtcgtattgctg 18
Db 305 atgcgcgtcgtattgctg 421

RESULT 8
AAZ12075
ID AAZ12075 standard; DNA; 1779 BP.
XX AC AAZ12075;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria gonorrhoeae complete ORF61 sequence.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea; ss.
XX OS Neisseria gonorrhoeae.
XX PN WO9924578-A2.
XX PD 20-MAY-1999.

```

```

XX PF 09-OCT-1998; 98WO-IB01665.
XX PR
XX PA (CHIR-) CHIRON SPA.
XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX PT WPI; 1999-327407/27.
XX DR P-PSDB; AAY38618.
XX PR
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX PS diagnosis, treatment and prevention of infection
XX PS Claim 9; Page 175; 524pp; English.
XX CC Nucleotide sequences AA211972-Z12358 represent open reading frames
XX CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
XX CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
XX CC fragments, their nucleic acids and antibodies are used for diagnosis,
XX CC prevention (as vaccines) or treatment of Neisseria infections,
XX CC such as meningitis, septicemia and gonorrhea. Both organisms
XX CC are closely related. Fragments of the nucleic acids are useful
XX CC as hybridisation probes and antisense reagents.
XX SQ Sequence 1779 BP; 367 A; 478 C; 599 G; 335 T; 0 other;

Query Match 85.6%; Score 15.4; DB 20; Length 1779;
Best Local Similarity 94.1%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgcgtattctg 18
||||| |||||||
Db 716 atgccgcgtattctg 732

RESULT 9
AAZ53673
ID AAZ53673 standard; DNA; 1779 BP.
XX AC
XX AC AAZ53673;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 311 partial DNA sequence SEQ ID NO:1205.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX KW antibacterial; gene therapy; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO9957280-A2.
XX PD
XX PF 11-NOV-1999.
XX PR 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.

25-FEB-1999; 99US-0121528.
XX (CHIR-) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizza M, Rappuoli R, Ratti C, Scalato E, Scarselli M;
XX PI Tettelin H, Venturi JC;
XX WPI; 2000-062150/05.
XX DR P-PSDB; AAY74911.
XX PT Novel Neisserial polyptides predicted to be useful antigens for
XX PT vaccines and diagnostics
XX PS Claim 7; Page 712-713; 1453pp; English.
XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polyptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polyptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polyptides can be used in the
XX CC manufacture of medicaments for treating or preventing infection due to
XX CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX CC presence of Neisseria bacteria, or to raise antibodies. They may also
XX CC be used to screen for agonists or antagonists, which may themselves
XX CC have use as antibacterial agents. The polynucleotides of the invention
XX CC may also be used in gene therapy protocols.
XX SQ Sequence 1779 BP; 367 A; 478 C; 599 G; 335 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 1779;
Best Local Similarity 94.1%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgcgtattctg 18
||||| |||||||
Db 716 atgccgcgtattctg 732

RESULT 10
AAZ14061
ID AAT14061 standard; DNA; 5859 BP.
XX AC
XX AC AAT14061;
XX DT 09-JUL-1996 (first entry)
XX DE N. gonorrhoeae lgt locus.
XX KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
XX KW vaccine; ds.
XX OS Neisseria gonorrhoeae strain F62.
XX FH Key Location/Qualifiers
XX FT CDS 445..1491
XX FT /*tag= a
XX FT /product= LgtA
XX FT misc_feature 699..715
XX FT /*tag= b
XX FT /note= "poly-C tract"
XX FT CDS 1491..2330
XX FT /*tag= c
XX FT /product= LgtB
XX FT CDS 2342..3262
XX FT /*tag= d
XX FT /product= LgtC
XX FT misc_feature 2499..2508
XX FT /*tag= e

```

FT CDS /note= "poly-G tract"
 FT 3322..4335
 FT /*tag= f
 FT /product= LgtD
 FT misc_feature 3576..3586
 FT /*tag= g
 FT /note= "poly-G tract"
 FT 4354..5196
 FT /*tag= h
 FT /product= LgtE

XX WO9610086-A1.

XX PD 04-APR-1996.

XX PF 25-SEP-1995; 95WO-US12317.

XX PR 26-SEP-1994; 94US-0312387.

XX PA (UYRQ) U:IV ROCKEFELLER.

XX PI Gotschlich EC;

XX WPI; 1996-200924/20.

XX DR P-PSDB; AAR91311, AAR91312, AAR91313, AAR91314, AAR91315.

XX Nucleic acids encoding glycosyl transferase(s) - used in the
 PT diagnosis of infection with Neisseria and for the biosynthesis of
 PT oligo:saccharide(s)

XX PS Claim 1; Fig 2g-m; 81pp; English.

XX CC The lgt locus (AAT140614) of Neisseria gonorrhoeae F62 contains
 CC 5 open reading frames, lgtA, lgtB, lgtC, lgtD and lgtE, coding for
 CC 5 glycosyltransferases (see also AAR91311-15) involved in gonococcal
 CC lipooligosaccharide (LOS) biosynthesis. The sequence was
 CC constructed from 2 clones isolated from an F62 gene bank in
 CC lambda-200 following screening with plasmid pRI0PI. 3 Of the
 CC coding sequences contain poly-G tracts that make them susceptible
 CC to premature termination. The lgt coding sequences can be used
 CC for prodn. of recombinant Lgt glycosyltransferases that are utilised
 CC in biosynthesis of LOS useful in vaccine prepn.

XX SQ Sequence 5859 BP; 1411 A; 1462 C; 1661 G; 1325 T; 0 other;

Query Match 85.6%; Score 15.4; DB 17; Length 5859;
 Best Local Similarity 94.1%; Prev. No. 54;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 atgcgcgtcgtattgctg 18

Db 2796 atgcgcgcgtattgctg 2812

RESULT 11

AAT49230

ID AAT49230 standard; DNA; 5859 BP.

XX AC AAT49230;

XX DT 21-MAR-1997 (first entry)

XX Lipo-oligosaccharide (including polyglycosyltransferase) gene.

XX Polyglycosyltransferase; N-acetylglucosaminyl transferase;

KW N-acetylgalactosaminyl transferase; lipo-oligosaccharide; ss.

XX OS Neisseria gonorrhoeae ATCC 33084.

XX FH Key Location/Qualifiers

XX CDS 1..381

FT /*tag= a

FT

FT CDS 445..1491
 FT /*tag= b
 FT /product= polyglycosyltransferase
 FT 2342..3262
 FT /*tag= c
 FT 3322..4335
 FT /*tag= d
 FT 4354..5196
 FT /*tag= e

XX PN WO9640971-A1.

XX PD 19-DEC-1996.

XX PF 03-JUN-1996; 96WJ-US08323.

XX PR 07-JUN-1995; 95US-0478140.

XX PA (NEOS-) NEOS TECHNOLOGIES INC.

XX PI Buczala SL, Johnson KF, Koth S;

XX WPI; 1997-052451/05.

XX DR P-PSDB; AAW06576, AAW06577, AAW06578, AAW06579, AAW06580.

XX PT Transfer of at least 2 saccharide units using

PT polyglycosyltransferase - isolated from N. gonorrhoeae, catalyses
 PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
 PT single galactose moiety

XX PS Disclosure; Fig 2A-B; 38pp; English.

XX CC A lipooligosaccharide-encoding gene region (AAT49230) of Neisseria
 CC gonorrhoeae ATCC 33084 includes a coding sequence for a novel N-
 CC polyglycosyltransferase (PGTase) (AAW06576) that catalyses the
 CC addition of both GlcNAc and GalNAc disaccharides to a single
 CC galactose moiety. The products (AAW06577-80) of the other coding
 CC sequences of this region are not identified. The PGTase gene can
 CC be obt. using standard techniques and incorporated into a vector
 CC to allow prodn. of the PGTase in transformed host cells. The
 CC enzyme is useful in the synthesis of oligosaccharides.

XX SQ Sequence 5859 BP; 1407 A; 1462 C; 1661 G; 1329 T; 0 other;

Query Match 85.6%; Score 15.4; DB 18; Length 5859;
 Best Local Similarity 94.1%; Prev. No. 54;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 atgcgcgtcgtattgctg 18

Db 2796 atgcgcgcgtattgctg 2812

RESULT 12

AAA53978/C

ID AAA53978 standard; DNA; 27425 BP.

XX AC AAA53978;

XX DT 08-FEB-2001 (first entry)

XX Dmp53 tumour suppressor gene (genomic DNA).

XX p53; tumour suppressor gene; insect; phenotype; metazoa;

KW invertebrate; screening; pharmaceutical; pesticide; mis-expression;

KW mutation; modulation; ds.

XX OS Drosophila melanogaster.

XX PN WO200055178-A1.

XX PD 21-SEP-2000.

```

XX 13-MAR-2000; 2000WO-US06602.
XX
XX 16-MAR-1999; 99US-0268969.
XX 23-FEB-2000; 2000US-0184373.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Buchman AR, Platt DM, Ollman MM, Young LM, Demsky MR, Keegan KP;
XX Friedman L, Koczynski C, Larson JS, Robertson SA;
XX
XX WPI; 2000-638178/61.
XX
XX Novel p53 tumor suppressor gene encoding a protein useful for
XX genetically modifying metazoan invertebrate organisms, such as insects
XX for screening compounds of pharmaceutical use or a pesticide
XX
XX Example 7; Page 80-89; 98pp; English.
XX
XX Insect p53 tumour suppressor genes can be used to genetically modify
XX metazoan invertebrate organisms, such as insects and worms, or
XX cultured cells, resulting in p53 expression or mis-expression. The
XX tumour suppressor genes, a p53 polypeptide or genetically modified
XX organisms or cells are used in screening assays to identify compounds
XX or molecules, preferably a pharmaceutical agent or a pesticide, that
XX modulates p53 activity. The genetically modified organisms or cells
XX are also useful for studying p53 activity by detecting the phenotype
XX caused by the expression or mis-expression of the p53 protein in the
XX insect. The method additionally comprises observing a second insect
XX having the same genetic modification as the previous one, which
XX causes the expression or mis-expression of the p53 protein, where the
XX second animal additionally comprises a mutation in a desired gene and
XX differences between the phenotype of the first and second identifies
XX the desired gene as capable of modifying the function of the gene
XX encoding the p53 protein. The genetically modified organisms or
XX cells are also useful for identifying other genes modulating the
XX function of, or interaction with the p53 gene.
XX
XX Sequence 27425 BP; 7309 A; 6492 C; 6305 G; 7319 T; 0 other;

Query Match      85.6%; Score 15.4; DB 21; Length 27425;
Best Local Similarity 94.1%; Pred. No. 60;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18
DB 9553 AGGCCGTCGTATTGCTG 9537

RESULT 13
AAQ20926/c
ID AAQ20926 standard; DNA; 742 BP.
XX
XX AAQ20926;
XX
XX 06-MAY-1992 (first entry)
XX
XX C10-E15 DNA fragment encoding NANBH-specific antigen polypeptide.
XX
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX
XX Non-A non-B hepatitis virus.
XX
XX Key Location/Qualifiers
XX CDS 3..742
XX /*tag= a
XX
XX EP468657-A.
XX
XX 29-JAN-1992.
XX
XX C10-E15 DNA fragment encoding NANBH-specific antigen polypeptide.
XX
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX
XX Non-A non-B hepatitis virus.
XX
XX Key Location/Qualifiers
XX CDS 3..742
XX /*tag= a
XX
XX EP468657-A.
XX
XX 29-JAN-1992.
XX
XX 08-JUL-1991; 91EP-0306158.

```

```

XX 20-DEC-1990; 90JP-0413844.
XX 09-JUL-1990; 90JP-0180889.
XX 30-NOV-1990; 90JP-0339589.
XX
XX (TOFU) TONEN CORP.
XX
XX Maki N, Yamaguchi K, Toyoshima A, Kohara M;
XX
XX WPI; 1992-034390/05.
XX P-PSDB; AAR20723.
XX
XX Non-A, non-B hepatitis-specific antigen polypeptide - for detection
XX of hepatitis virus gene or antibody directed against virus
XX
XX Claim 18; Fig 18; 78pp; English.
XX
XX The sequence is that of a DNA fragment encoding a non-A non-B (NANB)
XX hepatitis-specific antigen polypeptide, it can be used to produce the
XX polypeptide recombinantly. It is derived from the DNA clone C10-E15.
XX It can be used to give PCR primers which are capable of detecting NANB
XX hepatitis with high accuracy. See also AAQ20617-Q20629 and
XX AAQ20922-Q20925.
XX
XX Sequence 742 BP; 129 A; 217 C; 209 G; 187 T; 0 other;

Query Match      83.3%; Score 15; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgtattg 15
DB 372 CATGCCGTCGTATTG 358

RESULT 14
AAQ20923/c
ID AAQ20923 standard; DNA; 932 BP.
XX
XX AAQ20923;
XX
XX 06-MAY-1992 (first entry)
XX
XX C10-E12 DNA fragment encoding NANBH-specific antigen polypeptide.
XX
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX
XX Non-A non-B hepatitis virus.
XX
XX Key Location/Qualifiers
XX CDS 3..942
XX /*tag= a
XX
XX EP468657-A.
XX
XX 29-JAN-1992.
XX
XX 08-JUL-1991; 91EP-0306158.
XX
XX 20-DEC-1990; 90JP-0413844.
XX 09-JUL-1990; 90JP-0180889.
XX 30-NOV-1990; 90JP-0339589.
XX
XX (TOFU) TONEN CORP.
XX
XX Maki N, Yamaguchi K, Toyoshima A, Kohara M;
XX
XX WPI; 1992-034390/05.
XX P-PSDB; AAR20720.
XX
XX Non-A, non-B hepatitis-specific antigen polypeptide - for detection
XX of hepatitis virus gene or antibody directed against virus

```

XX PS Claim 15; Fig 15; 78pp; English.

XX CC The sequence is that of a DNA fragment encoding a non-A non-B (NANB)

XX CC hepatitis-specific antigen polypeptide, it can be used to produce the

XX CC polypeptide recombinantly. It is derived from the DNA clone C10-E12.

XX CC It can be used to give PCR primers which are capable of detecting NANB

XX CC hepatitis with high accuracy. See also AAQ20617-Q20629 and

XX CC AAQ20922-Q20926.

XX SQ Sequence 932 BP; 173 A; 276 C; 257 G; 226 T; 0 other;

Query Match 83.3%; Score 15; DB 13; Length 932;

Best Local Similarity 100.0%; Pred. No. 78;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtcgtattg 15

Db 438 CATGCCGTCGTATTG 424

RESULT 15

AAAX19484

ID AAX19484 standard; DNA; 1569 BP.

XX AC

XX AC AAX19484;

XX DT 21-MAY-1999 (first entry)

XX DE Streptococcus pneumoniae ffh encoding DNA.

XX KW Streptococcus pneumoniae; ffh; fifty-four homologue; antibacterial;

XX KW infection; otitis media; conjunctivitis; bacteraemia; sinusitis;

XX KW pleural empyema; endocarditis; meningitis; ds.

XX OS Streptococcus pneumoniae.

XX PN EP900843-A2

XX PD 10-MAR-1999.

XX PF 20-AUG-1998; 98EP-0306685.

XX PR 02-SEP-1997; 97US-0923772.

XX PA (SMK) SMITHLINE BEECHAM.

XX PI Black MT;

XX WPI; 1999-155936/14.

XX P-PSDB; AAW99658.

XX New Streptococcus pneumoniae Fifty-Four Homologue (Ffh) polypeptide

XX and polynucleotide - useful as diagnostic reagents and for

XX prevention and treatment of Streptococci infections, which cause

XX otitis media, sinusitis and conjunctivitis

XX Claim 4; Page 5-6; 21pp; English.

XX The present sequence encodes the Streptococcus pneumoniae fifty-four

XX homologue (ffh) protein, which is a component of the protein secretory

XX apparatus in bacteria, and the bacterial homologue of the eukaryotic

XX Signal Recognition Particle. Ffh proteins and polynucleotides are useful

XX for diagnosing diseases related to over or underexpression of Ffh protein or

XX by identifying mutations in the Ffh gene, or determining Ffh protein or

XX mRNA expression levels due to an infection of an organism with the Ffh

XX gene. They can diagnose the stage and type of infection. Ffh proteins are

XX also useful for screening for compounds which affect activity of the

XX protein by measuring the binding to ffh protein and observing the

XX stimulation or inhibition of the protein function. These can be used in

XX treatment to inhibit (antagonist i.e. antibacterial drugs) or enhance

XX (agonist) Ffh activity, in addition to direct administration of Ffh

CC proteins to treat conditions associated with a lack of Ffh protein, or

CC direct administration of antisense sequences to prevent expression. Ffh

CC proteins (administered directly, in a vector and as a vaccine) and

CC antibodies induce an immune response to immunise and prevent disease.

CC Diseases diagnosed, prevented or treated include: bacterial infections,

CC especially Streptococcus pneumoniae infections, which cause otitis media,

CC conjunctivitis, bacteraemia, sinusitis, pleural empyema, endocarditis and

CC especially meningitis. Ffh proteins, polynucleotides and their

CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and

CC are useful for use on wounds and body implants to prevent bacterial

CC infection.

XX SQ Sequence 1569 BP; 479 A; 297 C; 382 G; 411 T; 0 other;

Query Match 83.3%; Score 15; DB 20; Length 1569;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gccgcgcgtcgtattgctt 18

Db 1196 gccgcgcgtcgtattgctt 1210

Search completed: April 2, 2002, 22:49:19

Job time: 9219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 08:07:11 ; Search time 165.4 Seconds
(without alignments)
24.647 Million cell updates/sec

Title: US-09-785-904-4
Perfect score: 18
Sequence: 1 catgccgtcgattgctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	100.0	650	3	US-08-464-052-5
C 2	18	100.0	650	4	US-08-461-002-5
C 3	18	100.0	650	4	US-08-689-411-5
C 4	18	100.0	1535	3	US-08-464-052-1
C 5	18	100.0	1535	4	US-08-461-002-1
C 6	18	100.0	1535	4	US-08-689-411-1
C 7	18	100.0	1535	5	PCT-US94-09863-1
C 8	18	100.0	12412	1	US-08-390-878-18
C 9	18	100.0	4403765	4	US-09-103-840A-2
C 10	18	100.0	411529	4	US-09-103-840A-1
C 11	15.4	85.6	5859	1	US-08-312-387B-1
C 12	15.4	85.6	5859	1	US-08-312-387B-7
C 13	15.4	85.6	5859	1	US-08-683-426-1
C 14	15.4	85.6	5859	1	US-08-683-426-7
C 15	15.4	85.6	5859	1	US-08-683-458-1
C 16	15.4	85.6	5859	1	US-08-683-458-7
C 17	15.4	85.6	5859	1	US-08-878-360-1
C 18	15.4	85.6	5859	2	US-08-878-360-7
C 19	15.4	85.6	5859	3	US-08-478-140B-1
C 20	15	83.3	742	1	US-08-081-072-18
C 21	15	83.3	742	1	US-08-449-093A-18
C 22	15	83.3	932	1	US-08-081-072-15
C 23	15	83.3	932	1	US-08-449-093A-15
C 24	15	83.3	1569	2	US-08-923-772-1
C 25	14.8	82.2	576	1	US-08-086-428B-18
C 26	14.8	82.2	576	1	US-08-468-570-18
C 27	14.8	82.2	576	2	US-08-290-665A-18

C 28	14.8	82.2	576	5	PCT-US95-10398-18	Sequence 18, Appl
C 29	14.8	82.2	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 30	14.4	80.0	976	2	US-08-248-839C-5	Sequence 5, Appl
C 31	14.4	80.0	983	2	US-08-248-839C-1	Sequence 1, Appl
C 32	14.4	80.0	1200	2	US-08-248-839C-7	Sequence 7, Appl
C 33	14.4	80.0	4615	1	US-08-188-582-1	Sequence 1, Appl
C 34	14.4	80.0	4615	1	US-08-646-715-1	Sequence 1, Appl
C 35	13.8	76.7	272	1	US-08-362-670B-10	Sequence 10, Appl
C 36	13.8	76.7	272	1	US-08-362-670B-12	Sequence 12, Appl
C 37	13.8	76.7	272	3	US-08-333-576C-10	Sequence 10, Appl
C 38	13.8	76.7	272	3	US-08-333-576C-12	Sequence 12, Appl
C 39	13.8	76.7	272	4	US-08-808-324-10	Sequence 10, Appl
C 40	13.8	76.7	272	4	US-08-808-324-12	Sequence 12, Appl
C 41	13.8	76.7	272	5	PCT-US94-14030A-10	Sequence 10, Appl
C 42	13.8	76.7	272	5	PCT-US94-14030A-12	Sequence 12, Appl
C 43	13.8	76.7	926	1	US-08-362-670B-1	Sequence 1, Appl
C 44	13.8	76.7	926	3	US-08-333-576C-1	Sequence 1, Appl
C 45	13.8	76.7	926	4	US-08-808-324-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-464-052-5/c
; Sequence 5, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-5

Query Match 100.0%; Score 18; DB 3; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGATTGCTG 172

```
RESULT 2
US-08-461-002-5/c
; Sequence 5, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-5

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3
US-08-689-411-5/c
; Sequence 5, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 4
US-08-464-052-1/c
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-464-052-1

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 5
US-08-461-002-5
; Sequence 5, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-5

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 6
US-08-689-411-5/c
; Sequence 5, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```


Query Match 100.0%; Score 18; DB 3; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtctgtattgctg 18
|||||
Db 1074 CATGCCGTCTGTATTGCTG 1057

RESULT 5

US-08-461-002-1/c
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

Query Match 100.0%; Score 18; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtctgtattgctg 18
|||||
Db 1074 CATGCCGTCTGTATTGCTG 1057

RESULT 6

US-08-689-411-1/c
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: US/08/689,411
CLASSIFICATION: 516
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/187
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-689-411-1

Query Match 100.0%; Score 18; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtctgtattgctg 18
|||||
Db 1074 CATGCCGTCTGTATTGCTG 1057

RESULT 7

PCT-US94-09863-1/c
; Sequence 1, Application PC/TUS9409863
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
; UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael L. Goldman
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: PCT/US94/09863
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman Mr., Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1000
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1

Query Match 100.0%; Score 18; DB 5; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgctattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 8
US-08-390-878-18/c
; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; CITY: San Francisco
; STATE: California --
; COUNTRY: USA
; ZIP: 94105

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 100.0%; Score 18; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgctattgctg 18
|||||
Db 705 CATGCCGTCGTATTGCTG 688

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENG.H: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: (DNC) 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, L, c or g
US-09-103-840A-2

Query Match 100.0%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgctattgctg 18
|||||
Db 104467 catgccgtcgctattgctg 104484

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 18; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgctattgctg 18
|||||
Db 104479 catgccgtcgctattgctg 104496

RESULT 11
US-08-312-387B-1
; Sequence 1, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
222

;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/683,426
;; FILING DATE:
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/312,387
;; FILING DATE: September 26, 1994
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-095B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5859 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Neisseria gonorrhoeae
;; STRAIN: F62
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..381
;; OTHER INFORMATION: glys (glycyl tRNA synthetase beta chain)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 445..1491
;; OTHER INFORMATION: lgtA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2342..3262
;; OTHER INFORMATION: lgtC
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3322..4335
;; OTHER INFORMATION: lgtD
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 4354..5196
;; OTHER INFORMATION: lgtE
;;
;; US-08-683-426-1

Query Match 85.6%; Score 15.4; DH 1; Length 5859;
Best Local Similarity 94.1%; Pred. No. 19;
Matches 1f; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18
||||| |||||||||

Db 2796 ATGCCGGCGTATTGCTG 2812

RESULT 14
US-08-683-426-7
; Sequence 7, Application US/08683426
; Patent No. 5703367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack

;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/683,426
;; FILING DATE:
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/312,387
;; FILING DATE: September 26, 1994
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-095B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5859 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Neisseria gonorrhoeae
;; STRAIN: F62
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1491..2330
;; OTHER INFORMATION: lgtB
;;
;; US-08-683-426-7

Query Match 85.6%; Score 15.4; DB 1; Length 5859;
Best Local Similarity 94.1%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18
||||| |||||||||

Db 2796 ATGCCGGCGTATTGCTG 2812

RESULT 15
US-08-683-458-1
; Sequence 1, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25

Wed Apr 3 08:33:28 2002

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: F62
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..381
; OTHER INFORMATION: glycyl tRNA syntetase beta chain)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 445..1491
; OTHER INFORMATION: lgtA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2342..3262
; OTHER INFORMATION: lgtC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3322..4335
; OTHER INFORMATION: lgtD
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4354..5196
; OTHER INFORMATION: lgtE
;
; US-08-683-458-1

```

```

Query Match      85.6%; Score 15.4; DB 1; Length 5859;
Best Local Similarity 94.1%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 atgccgcgtattgctg 18
        ||||| |||||
Db      2796 ATGCCGCGTATGCTG 2812

```

Search completed: April 3, 2002, 08:12:53
Job time: 42533 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 21:48:01 ; Search time 3835.7 Seconds
(without alignments)
50.427 Million cell updates/sec

Title: US-09-785-904-4

Perfect score: 18
Sequence: 1 catgcgcgtatgctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_esthr: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_est1: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.4	91.1	779	13 CNS02LR5	AL203162 Tetraodon
C 2	16.4	91.1	1038	13 CNS02KIX	AL201570 Tetraodon
C 3	15.4	85.6	252	10 AU060300	AU060300 Tetraodon
C 4	15.4	85.6	332	10 AU061778	AU061778 Tetraodon
C 5	15.4	85.6	468	11 N93955	N93955 zb74f01.s1
C 6	15.4	85.6	496	10 AU061934	AU061934 Tetraodon
C 7	15.4	85.6	508	13 AQ179545	AQ179545 HS_3177_A
C 8	15.4	85.6	529	10 BE775420	BE775420 MY-02-G-0
C 9	15.4	85.6	530	10 AW934449	AW934449 EST340426
C 10	15.4	85.6	640	10 AI403760	AI403760 GH23265.5
C 11	15.4	85.6	671	10 BE777156	BE777156 MY-25-B-0
C 12	15.4	85.6	2051	11 BF868166	BF868166 963101A02

13	15	83.3	377	13	BH100992	BH100992 RPI-24-3
14	15	83.3	575	10	AV834783	AV834783 AV834783
15	14.8	82.2	128	10	A1105888	A1105888 cb01a01_p
C 16	14.8	82.2	315	13	B28664	B28664 T15K18TF_TA
C 17	14.8	82.2	383	11	BG979425	BG979425 CM4-CN009
C 18	14.8	82.2	391	13	A0085253	A0085253 HS_2163_A
C 19	14.8	82.2	395	10	BE428498	BE428498 MT0007.F1
C 20	14.8	82.2	410	11	R41436	R41436 Yf95d12.s1
C 21	14.8	82.2	425	13	BE442482	BE442482 925021C06
C 22	14.8	82.2	428	13	A0646244	A0646244 RPI193-EC
C 23	14.8	82.2	434	10	BE024649	BE024649 894004F08
C 24	14.8	82.2	464	10	BE024887	BE024887 894006F11
C 25	14.8	82.2	479	11	BF848907	BF848907 QVI-EN005
C 26	14.8	82.2	486	10	AA314278	AA314278 EST186121
C 27	14.8	82.2	493	11	BG733499	BG733499 347421_MA
C 28	14.8	82.2	505	11	H11823	H11823 Ym11905.r1
C 29	14.8	82.2	523	11	BF769545	BF769545 KC3-IT001
C 30	14.8	82.2	534	13	TA72803P	TA72803P T. brucei
C 31	14.8	82.2	534	13	A0555508	A0555508 HS_5222_A
C 32	14.8	82.2	548	11	BT343257	BT343257 371350_MA
C 33	14.8	82.2	604	13	A0370996	A0370996 HS_5034_B
C 34	14.8	82.2	640	10	AW954845	AW954845 EST366915
C 35	14.8	82.2	660	11	BG858511	BG858511 1024058B0
C 36	14.8	82.2	666	13	AZ829376	AZ829376 2M0107F04
C 37	14.8	82.2	696	10	AU134159	AU134159 AU134159
C 38	14.8	82.2	701	11	BG860192	BG860192 1024068A0
C 39	14.8	82.2	714	10	AU213234	AU213234 AU213234
C 40	14.8	82.2	714	13	BH110072	BH110072 RPI-24-3
C 41	14.8	82.2	742	13	AQ250765	AQ250765 T15M3-T7
C 42	14.8	82.2	780	11	BF942570	BF942570 6D4_CDNA
C 43	14.8	82.2	790	10	AW940855	AW940855 GH22341.3
C 44	14.8	82.2	844	13	AZ129739	AZ129739 OSJNB0006
C 45	14.8	82.2	924	11	BG028677	BG028677 602292844

ALIGNMENTS

RESULT 1
CNS02LR5/c
LOCUS
DEFINITION

CNS02LR5 779 bp DNA GSS 14-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
148A13 of library G from Tetraodon nigroviridis, genomic survey
sequence.

AL203162

AL203162.1 G1:7861507

GSS: genome survey sequence.

KEYWORDS Tetraodon nigroviridis.

SOURCE Tetraodon nigroviridis

ORGANISM Tetraodon nigroviridis

Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 779)

AUTHORS

Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 779)

AUTHORS

Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 779)

AUTHORS

Genoscope.

Direct Submission

TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES

source
Location/Qualifiers

1..779
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/note="Genoscope sequence ID : COAG148AA07Lp1-end : T7"

BASE COUNT 187 a 178 c 268 g 142 t 4 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 779;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18

Db 139 CATGCCGTCGTATTCCTG 122

RESULT

CNS02KIX 1038 bp DNA GSS 14-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 14SH15 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL201570
VERSION AL201570.1 GI:7859915
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1038)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1038)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished
AUTHORS 3 (bases 1 to 1038)
Genoscope.
TITLE Direct Submission
COMMENT Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source
Location/Qualifiers

1..1038
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/note="Genoscope sequence ID : COAG145CD08Lp1-end : T7"

BASE COUNT 240 a 291 c 255 g 243 t 9 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 1038;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18

Db 885 CATGCCGTCGTATTCCTG 902

RESULT

AU060300/c 252 bp mRNA EST 20-MAY-1999
LOCUS Tetraodon nigroviridis discoideum SL (H.Urushihara) Dictyostelium
DEFINITION discoideum cDNA clone SLJ109, mRNA sequence.

ACCESSION AU060300
VERSION AU060300.1 GI:4881404

KEYWORDS EST.
SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 252)
Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara

INSTITUTE OF Biological Sciences
UNIVERSITY OF Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@tsukuba.cc.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES

source
Location/Qualifiers

1..252
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLJ109"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 73 a 62 c 41 g 76 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 252;
Best Local Similarity 94.1%; Pred. No. 9.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgtcgattgctg 18

Db 83 AAGCCGTCGTATTCCTG 57

RESULT

AU061778/c 332 bp mRNA EST 20-MAY-1999
LOCUS Tetraodon nigroviridis discoideum SL (H.Urushihara) Dictyostelium
DEFINITION discoideum cDNA clone SLF846, mRNA sequence.

ACCESSION AU061778
VERSION AU061778.1 GI:4882882

KEYWORDS EST.
SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 332)
Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara

INSTITUTE OF Biological Sciences
UNIVERSITY OF Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@tsukuba.cc.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES

source

Location/Qualifiers

1. .332
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLF846"
 /dev_stage="slug"
 /dev_stage="slug"

BASE COUNT 94 a 92 c 49 g 97 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 332;
 Best Local Similarity 94.1%; Pred. No. 9.6e-02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcccgtctattgctg 18

|||||

Db 83 AAGCCGTCGTATTGCTG 67

RESULT 5

N93955/c

LOCUS

zbf4f01.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
 IMAGE:309337 3' similar to gb:M64497 APOLIPOPROTEIN A1 REGULATORY
 PROTEIN-1 (HUMAN); mRNA sequence.

ACCESSION

N93955

VERSION

N93955.1 GI:1266264

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 468)

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins

, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore

, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,

Schellenberg, K., Soares, J.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,

Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 480 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 223.

Location/Qualifiers

1. .468

/organism="Homo sapiens"

/db_xref="GDB:1252750"

/db_xref="taxon:9606"

/clone="IMAGE:309337"

/clone_lib="Soares_fetal_lung_NbHL19W"

/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: lung; Vector: p7T3D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGCGGCGCAATTTTTTTTTTTT-3']

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p7T3 vector

(Pharmacia). Library went through one round of

normalization to a Cot ~ 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NBHL19W."

BASE COUNT 118 a 102 c 74 g 168 t 6 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 468;

Best Local Similarity 88.9%; Pred. No. 1e+03;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 catgcgcgtctattgctg 18

|||||

Db 383 CATNCCGTCGTATTGCTG 366

RESULT 6

AU061934/c

LOCUS

AU061934

DEFINITION

Dictyostellium discoideum SL (H.Urushihara) Dictyostellium

discoideum cDNA clone SLG680, mRNA sequence.

ACCESSION

AU061934

VERSION

AU061934.1 GI:4883038

KEYWORDS

EST.

SOURCE

Dictyostellium discoideum.

ORGANISM

Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.

REFERENCE

AUTHORS

Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,

Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,

Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

Developmental cDNA in Dictyostellium discoideum

Unpublished (1998)

Contact: Hiteko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-ndai, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = Dictyostellium discoideum cDNA project in Japan.

Location/Qualifiers

1. .496

/organism="Dictyostellium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SLG680"

/clone_lib="Dictyostellium discoideum SL (H.Urushihara)"

/dev_stage="Slug"

146 a 119 c 79 g 152 t

BASE COUNT

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 496;

Best Local Similarity 94.1%; Pred. No. 1e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcccgtctattgctg 18

|||||

Db 83 AAGCCGTCGTATTGCTG 67

RESULT 7

AQ179545

LOCUS

AQ179545

DEFINITION

HS_1177_A1_E04_MR ClT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=3177 Col=7 Row=I, DNA sequence.

ACCESSION

AQ179545

VERSION

AQ179545.1 GI:3576912

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 508)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3177 row: I column: 7
Class: BAC ends
High quality sequence stop: 508.

FEATURES source
1..508
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate=3177 Col=7 Row=I
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 169 a 115 c 46 g 175 t 3 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 13; Length 508;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgcgtatttgcgt 17

|||||

Db 456 CATGCCGTCGTATTACT 472

RESULT 8

LOCUS BE775420 529 bp mRNA EST 20-SEP-2000

DEFINITION MY-02-G-06 pinfestansMY Phytophthora infestans cDNA, mRNA sequence.

ACCESSION BE775420

VERSION BE775420.1 GI:10229130

KEYWORDS EST.

SOURCE potato late blight agent.

ORGANISM Phytophthora infestans

Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;

Phytophthora.

1 (bases 1 to 529)

Kanoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.

Initial assessment of gene diversity for the oomycete pathogen

Phytophthora infestans based on expressed sequences

Fungal Genet. Biol. 28 (2), 94-106 (1999)

JOURNAL 20056376

MEDLINE Contact: Govers F

COMMENT Laboratory of Phytopathology

Wageningen University

Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands

Tel: 31 317 483 138

Fax: 31 317 483 412

Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES source

1..529
Location/Qualifiers

/organism="Phytophthora infestans"

/strain="DDR7602, A1 mating type"

/db_xref="taxon:4787"

/clone_lib="pinfestansMY"

/dev_stage="4-week old vegetative, non-sporulating

mycelium in synthetic medium"

/lab_host="E. coli, strain DH5-alpha"

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total RNA was isolated from mycelium of P. infestans DDR7602 cultured for 4 weeks in synthetic medium. EST clones were named by their position in the microtiter plate, preceded by the prefix MY (for mycelium) and the successive number of the microtiter plate (e.g. MY-06-A-04)."

BASE COUNT 116 a 155 c 124 g 107 t 22 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 529;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgcgtatttgcgt 17

|||||

Db 318 CATGCCGTCGTATTGCT 302

RESULT 9

LOCUS AW934449/c

DEFINITION EST340426 tomato fruit mature green, TAMU Lycopersicon esculentum

ACCESSION AW934449

VERSION AW934449.1 GI:8105370

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 530)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,

Upton,J., Hansen,T., Craven,M.B., Rowman,C.L., Ahn,S., Kanning,C.M.,

Fraser,C.M., Martin,C.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato fruit tissue

Unpublished (1999)

Contact: CUCI

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES Location/Qualifiers

1..530

/organism "Lycopersicon esculentum"

/cultivar "FAL9"

/db_xref="taxon:4081"

/clone_lib="tomato fruit mature green, TAMU"

/issue_type="fruit pericarp"

/dev_stage="mature green (3-5 days pre-ripening)"

/lab_host="SOLK"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XbaI; cDEF - fruit were tagged at the 1cm stage and

harvested 3-5 days prior to ripening. Fruit were cut in

half to verify the seeds were indeed 'immature' and the

seeds and locules were discarded prior to freezing the

pericarp"

BASE COUNT 164 a 117 c 103 g 146 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 530;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgcgtatttgcgt 18

|||||

Db 208 ATACCGTCGTATTGCTG 192

```

RESULT 10
A1403760
LOCUS GH23265.5prime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION melanogaster cDNA clone GH23265 5prime, mRNA sequence.
ACCESSION A1403760
VERSION A1403760.1 GI:4246847
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 640)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
High quality sequence stop: 438.
Location/Qualifiers
1..640
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH23265"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 116 a 192 c 204 g 128 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 640;
Best Local Similarity 94.1%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcgcgtcgattgctg 18
|||||
Db 530 AGCGCGTGGTATTGCTG 546

RESULT 11
BE777156
LOCUS MY-25-B-07 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
DEFINITION BE777156
ACCESSION BE777156
VERSION BE777156.1 GI:10230811
KEYWORDS EST.
SOURCE potato late blight agent.
ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
REFERENCE 1 (bases 1 to 671)
AUTHORS Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
TITLE Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
JOURNAL Fungal Genet. Biol. 28 (2), 94-106 (1999)
MEDLINE 20056376
COMMENT Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands

Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.
Location/Qualifiers
1..671
/organism="Phytophthora infestans"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/clone_lib="PinfestansMY"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DH5-alpha"
/note="Vector: pSPOKT1; Site_1: SalI; Site_2: NotI; Total
cDNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."
BASE COUNT 155 a 171 c 186 g 157 t 2 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 671;
Best Local Similarity 94.1%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcgcgtcgattgctg 18
|||||
Db 429 ATTCGCGTGGTATTGCTG 345

RESULT 12
BF868166/c
LOCUS Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
DEFINITION BF868166
ACCESSION BF868166
VERSION BF868166.1 GI:12258310
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 2051)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauscer,C.,
Lefebvre,F., McElmud,J.P., Shrago,J., Sillow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Back to
Unicellular System for Studying Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauscer
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: claus@duke.edu.
Location/Qualifiers
1..2051
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr). TAP N (40 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (40min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
```

pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 578 a 461 c 551 g 294 t 167 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 2051;
Best Local Similarity 88.9%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 catgcgtcgtattgctg 18
|||||

DB 1664 CATGCCGTCGTTCGCTG 1647

RESULT 13
BH100992
LOCUS
DEFINITION RP-CI-24-333F5.TVB RPCI-24 Mus musculus genomic clone RPCI-24-333F5, DNA sequence. GSS 19-JUL-2001
ACCESSION BH100992
VERSION BH100992.1 GI:14925413
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E., Russell,D., de-Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-333F5.TJ

CONTACT: Shaving Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from HACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end plate: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..377
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-333F5"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 81 a 65 c 78 g 153 t

ORIGIN

Query Match 83.3%; Score 15; DB 13; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgccgtcgtattgct 17
|||||

DB 4 TGCCGTCGTATTGCT 18

RESULT 14
AV834783
LOCUS
DEFINITION AV834783 K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage Hordeum vulgare subsp. spontaneum cDNA clone bahl5n23, mRNA sequence. EST 22-JUN-2001
ACCESSION AV834783
VERSION AV834783.1 GI:14526872
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. spontaneum.
ORGANISM Hordeum vulgare subsp. spontaneum

REFERENCE
AUTHORS Sato,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2001)
COMMENT Contact: Kazuhiro Sato
Research Institute for Bioresources
Okayama University, Barley Germplasm Center
Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
Email: kassato@rib.okayama-u.ac.jp,
URL: http://www.rib.okayama-u.ac.jp/barley/
Sato,K., Saisho,D., Takeda,K., Shini,T. and Kohara,Y. Direct
submission:
database: http://www.shigen.nig.ac.jp/barley/Barley.html.
Location/Qualifiers
1..575

FEATURES
source
1..575
/organism="Hordeum vulgare subsp. spontaneum"
/cultivar="H602"
/db_xref="taxon:77009"
/clone="bahl5n23"
/clone_lib="K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage"
/dev_stage="adult, heading stage"
/lissue_type="top three leaves"
BASE COUNT 142 a 160 c 154 g 119 t
ORIGIN

Query Match 83.4%; Score 15; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gccgtcgtattgctg 18
|||||

DB 503 GCCGTCGTATTGCTG 517

RESULT 15
A1105888
LOCUS
DEFINITION A1105888 128 bp mRNA EST 25-AUG-1998
cb01a01_p1 ZF adult heart library Danio rerio cDNA 5 prime, mRNA sequence.
ACCESSION A1105888
VERSION A1105888.1 GI:3460991
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 128)
Chen,J.N., deSauvage,F., Hosobuchi,M., Jackson,D.G. and Fishman,M.C.
TITLE Expressed Sequences from The Adult zebrafish Heart

Wed Apr 3 08:33:30 2002

JOURNAL Unpublished (1998)
COMMENT Contact: Mark C. Fishman
Cardiovascular Research Center
Massachusetts General Hospital
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
Fax: 6177265806
Email: fishman@cvc.harvard.edu
http://zebrafish.mgh.harvard.edu
The original clones used for sequencing are no longer available;
the library is available from Mark C. Fishman.
Insert length: 128 Std Error: 0.00
Seq primer: p1.

FEATURES Location/Qualifiers
source
1..128
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="Zf adult heart library"
/sex="mixed"
/tissue_type="myocardium, endocardium, vessel"
/dev_stage="adult"
/lab_host="E. coli XL1 Blue"
/note="Organ: heart; Vector: LambdaZAPII; Site_1: EcoRI;
Site_2: XhoI"
23 a 36 c 29 g 31 t 9 others

BASE COUNT
ORIGIN

Query Match 82.2%; Score 14.8; DB 10; Length 128;
Best Local Similarity 88.9%; Pred. No. 1.6e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 catgccgtcgtattgctg-18"
||| |
Db 33 CAAGCGTCGTATTGCTG 50

Search completed: April 2, 2002, 21:48:05
Job time: 10830 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:49:26 ; Search time 1472.17 Seconds
(without alignments)
5076.331 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccgggtgcgcgcgcatc 453

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	176	38.9	650	6	AR096715	AR096715 Sequence
2	176	38.9	650	6	AR147696	AR147696 Sequence
3	176	38.9	1535	6	AR096713	AR096713 Sequence
4	176	38.9	1535	6	AR147694	AR147694 Sequence
5	176	38.9	1604	1	MH035021	U35021 Mycobacteri
6	176	38.9	1453	1	MH043540	U3540 Mycobacteri
7	176	38.9	9281	1	MH043540	U35017 Mycobacteri
8	176	38.9	9764	1	AE006921	AE006921 Mycobacte
9	176	38.9	10019	1	AF041819	AF041819 Mycobacte
10	176	38.9	12412	6	186264	186264 Sequence 18
11	176	38.9	17783	1	AE007028	AE007028 Mycobacte
12	176	38.9	32437	1	MTCY336	295586 Mycobacteri
13	176	38.9	38380	1	MTCY336	274410 Mycobacteri
14	176	38.9	43401	1	MTCY336	295330 Mycobacteri
15	165	36.4	16384	1	AE007160	AE007160 Mycobacte
16	27	6.0	14179	1	AE007053	AE007053 Mycobacte
17	27	6.0	33100	1	MTCY9F9	284498 Mycobacteri
18	20	4.4	14408	1	AE005028	AE005028 Halobacte
19	20	4.4	17712	1	AE004844	AE004844 Pseudomon
20	19	4.2	9793	1	AE004944	AE004944 Pseudomon
21	19	4.2	1486	1	AE004770	AE004770 Pseudomon
22	19	4.2	28826	1	MTCY041	AE021958 Mycobacte
23	19	4.2	24606	1	U70376	U70376 Streptomyce
24	19	4.2	99699	2	AC013845	AC013845 Prosophi
25	19	4.2	126340	2	DMB18C8	AL122030 Prosophi
26	19	4.2	150621	2	AP003270	AP003270 Cryza sat
27	19	4.2	151490	8	AP003611	AP003611 Cryza sat
28	19	4.2	152237	8	AP001073	AE003438 Prosophi
29	19	4.2	299537	3	AE003438	AP003002 Mesorhizo
30	19	4.2	449498	1	AF003002	AF116759 Homo sapi
31	18	4.0	508	9	F116756S04	AF1245508 Colts go
32	18	4.0	521	5	CG0245508	AC057017 Giardia i
33	18	4.0	708	2	AC057017	AC077697 Giardia i
34	18	4.0	944	2	AC077697	AC030797 Giardia i
35	18	4.0	951	2	AC030797	AC083707 Giardia i
36	18	4.0	1002	2	AC030797	AF124531 Homo sapi
37	18	4.0	1193	9	HSP63G04	AX000494 Sequence
38	18	4.0	1206	6	AX000494	AX064129 Sequence
39	18	4.0	1413	6	AX064129	AX120935 Sequence
40	18	4.0	1413	6	AX120935	AX028480 Sequence
41	18	4.0	1536	6	AX028480	M12799 Corynebacte
42	18	4.0	1639	1	C000000R	AF160742 Corynebact
43	18	4.0	1701	1	AF160742	AF114243 Corynebact
44	18	4.0	1852	1	AF114243	E00311 DNA coding
45	18	4.0	1853	6	E00311	

ALIGNMENTS

RESULT 1	AR096715	AR096715	650 bp	INA	PAT	08-SEP-2000
LOCUS	Sequence	5 from patent US 6008201.				
DEFINITION	AR096715					
ACCESSION	AR096715.1	GI:10025753				
VERSION						
KEYWORDS	Unknown.					
SOURCE	Unclassified.					
ORGANISM	1 (bases 1 to 650)					
REFERENCE	DNA molecule encoding for cellular uptake of mycobacterium					
AUTHORS	Riley, L.W.					
TITLE	tuberculosis and uses thereof					
JOURNAL	Patent: US 6008201-A 5 28-DEC-1999;					
FEATURES	Location/Qualifiers					
source	1..650					
BASE COUNT	123 a 258 c 185 g 84 t					
ORIGIN						

FEATURES		location/Qualifiers		source			
		1..1535		/organism="unknown"			
BASE COUNT		297 a		544 c 458 g			
ORIGIN		236 t		0;			
Query Match							
Best Local Similarity 38.9%; Score 176; DB 6; Length 1535;							
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Pred. No. 4.2e-78;							
Qy	278	cgcgggccacatttgaagccgtctagcaaaactggccgcccccgagcgagcaaacccg 337					
Db	1106	CGCGGGCCACCTTTGAAGCCGTGCTAGCCAAATGGCCGCCCGCCGCCGACCAACCCG 1165					
Qy	338	acgaccacaccccggtcatcgacacaccccccgcgatggcgcccatcgaccgcgacacc 397					
Db	1166	ACGACCACACCCCGTCTATGACACACACCCCGATGCGCGGCATCGACCGGACACCC 1225					
Qy	398	gcagccaaagcccaacgcaacacagcggtgctggcgggctggcgcgctgctgac 453					
Db	1226	GCAGCAAGCCCAACGCAACACACGCGGTGCTGGCGGGTGGCGGCGCTGATC 1281					
RESULT 4							
LOCUS		ARI147694		1535 bp DNA			
DEFINITION		Sequence 1 from patent US 6224881.		PAT			
ACCESSION		ARI147694		08-AUG-2001			
VERSION		ARI147694.1		GI:15111784			
KEYWORDS		Unknown.					
SOURCE		Unknown.					
ORGANISM		Unclassified.					
REFERENCE		1 (bases 1 to 1535)					
AUTHORS		Riley,L.W. and Chong,P.					
TITLE		DNA molecule fragments encoding for cellular uptake of					
JOURNAL		Mycobacterium tuberculosis and uses thereof					
FEATURES		Patent: US 6224881-A 1 01-MAY-2001;					
Source		Location/Qualifiers					
BASE COUNT		297 a		544 c 458 g			
ORIGIN		236 t		0;			
Query Match							
Best Local Similarity 38.9%; Score 176; DB 6; Length 1545;							
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Pred. No. 4.2e-78;							
Qy	278	cgcgggccacatttgaagccgtctagcaaaactggccgcccccgagcgagcaaacccg 337					
Db	1106	CGCGGGCCACCTTTGAAGCCGTGCTAGCCAAATGGCCGCCCGCCGCCGACCAACCCG 1165					
Qy	338	acgaccacaccccggtcatcgacacaccccccgcgatggcgcccatcgaccgcgacacc 397					
Db	1166	ACGACCACACCCCGTCTATGACACACACCCCGATGCGCGGCATCGACCGGACACCC 1225					
Qy	398	gcagccaaagcccaacgcaacacagcggtgctggcgggctggcgcgctgctgac 453					
Db	1226	GCAGCAAGCCCAACGCAACACACGCGGTGCTGGCGGGTGGCGGCGCTGATC 1281					
RESULT 5							
LOCUS		MBU35021		1604 bp DNA			
DEFINITION		Mycobacterium bovis BCG DNA flanking deletion region 3.		BCI			
ACCESSION		U35021		27-APR-1996			
VERSION		U35021.1		GI:1049243			
KEYWORDS		Mycobacterium bovis BCG.					
SOURCE		Mycobacterium bovis					
ORGANISM		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					

FEATURES		location/Qualifiers		source			
		1..1535		/organism="unknown"			
BASE COUNT		297 a		544 c 458 g			
ORIGIN		236 t		0;			
Query Match							
Best Local Similarity 38.9%; Score 176; DB 6; Length 1545;							
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Pred. No. 4.2e-78;							
Qy	278	cgcgggccacatttgaagccgtctagcaaaactggccgcccccgagcgagcaaacccg 337					
Db	1106	CGCGGGCCACCTTTGAAGCCGTGCTAGCCAAATGGCCGCCCGCCGCCGACCAACCCG 1165					
Qy	338	acgaccacaccccggtcatcgacacaccccccgcgatggcgcccatcgaccgcgacacc 397					
Db	1166	ACGACCACACCCCGTCTATGACACACACCCCGATGCGCGGCATCGACCGGACACCC 1225					
Qy	398	gcagccaaagcccaacgcaacacagcggtgctggcgggctggcgcgctgctgac 453					
Db	1226	GCAGCAAGCCCAACGCAACACACGCGGTGCTGGCGGGTGGCGGCGCTGATC 1281					
RESULT 5							
LOCUS		MBU35021		1604 bp DNA			
DEFINITION		Mycobacterium bovis BCG DNA flanking deletion region 3.		BCI			
ACCESSION		U35021		27-APR-1996			
VERSION		U35021.1		GI:1049243			
KEYWORDS		Mycobacterium bovis BCG.					
SOURCE		Mycobacterium bovis					
ORGANISM		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					

[illegible]


```
/db_xref="taxon:33892"
complement(259..1038)
/note="ORF1: similar to MTCY336.37"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96954.1"
/db_xref="GI:2791843"
/translation="MKRSMKSGPAIGLAMLPVMAAPGLAAADPATRPVDYQQITD
VYIARGLSORGVPSWAGGIGSPTRGTGTINTVGFDAAGLIQYAYAGAGLKLPRSS
COMKYGOKVLPQOARKGDLIFVCEPTQSVALLYLKGOMLEVDGVVQVQSPVTRNGMT
PYLVFGSPSRPNSRRKSSQRRSSKRPSSKRGARPTGAGTGAAGTGAAGTGAAGTGA
PAASLRHALTITVPLAFSPKPAFCPLLSARSRLVWVN"
complement(1278..1850)
/note="ORF2: similar to MTCY336.36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96955.1"
/db_xref="GI:2791844"
/translation="MICSQVNPVVAAMEIEGNAATPRPIPVTVAKTSAIGEHACMAP
PNOPEDTQDAGTSSRSALDEPDBTTPRVLTIRIAPWDVVCVTVIILAVLVMVTMS
WPSRLFAFTDNVCPDACPVPFGVNYIYIPVMWGGIGCAALATATVIGPFVSMKGGWM
SWPPIISTAVITVTSIAGYATAGFSERYWH"
1810..3123
/genes="bioA"
/genes="bioA"
/genes="bioA"
/genes="bioA"
/genes="BioA: similar to MTCY336.35c"
/codon_start=1
/transl_table=11
/product="putative
adenosylmethionine-8-amino-7-oxononanoate
aminotransferase"
/protein_id="AAB96956.1"
/db_xref="GI:2791845"
/translation="NAAATGGTLTPEQIIAIVDGAHLWHPYSSIGKEAVSPVVAIIIGA
WLTLRDQGPVLELDAMSSWMTA IHGHGHPALDQALTOTLKVMIHNVFPGGLTTPHFAAR
LAKLIVITPAGLDIVFSDSGSVSEVAAKMAYGRGLPCKRRLMWRGCVHGD
TFLAMISCPHGMMSLMDVLAQAQVFAPOVPRDYPAYSAFAQLAQHAGELAAVV
VEPVQAGMRPHDPYRLDHLDRICRRREVLLIIDEIATGFGRTGALFAADHAGVSP
DINCVKATGGYLSLAATLTCTADVAHTISAAGALMGHPTFMANPLACAVSVASVE
LLLGQDMRTTRITELAAAGTAGLDTRALPAVTDVRCVGAIGVIECDRPVDVLAVAPAA
LDRGVYLRPRFNLVAMPYICTPAEITQITSAMVEVARLVGLP"
3120..4280
/genes="bioF"
/genes="bioF"
/genes="bioF"
/EC_number="2.3.1.47"
/note="BioF: similar to MTCY336.34c"
/codon_start=1
/transl_table=11
/product="putative 8-amino-7-oxononanoate synthase"
/protein_id="AAB96957.1"
/db_xref="GI:2791846"
/translation="MKAATQARIDDSPLAWLDAVQORHEAGIRKCLRPRAVATELD
LAGNDYLGLSRHPAVIDGGVQALRIWAGATGSRIVGTDKLHQOFEAELEAFVGA
GLFSSGYTANIGAVVGLSGPGLLVSDARSHASLVDACRLSRARVVVTPHRDVAVD
AALSRDEOPAVVTDVSFSDGSLAPVRELLEVCRRHALLVDEAHGIGVVRGGGRG
LLYELGAGAPDVVWTTLSKALGSGGVVLTGPVVRHLIDAARPFIFDTGLAPAAV
GAARALRLQAEPRPQAVLNVHAGELARMCGVAAPVDSAMYSVILGEPSAVAAAA
CLDAGYKVCGRFPPTVPAGTSRLRLTARASNLAGELELARRVLTDLAVARR"
4277..4957
/genes="bioD"
/genes="bioD"
/genes="bioD"
/EC_number="6.3.3.3"
/note="BioD: similar to MTCY336.33c"
/codon_start=1
/transl_table=11
/product="putative dethiobiotin synthetase"
/protein_id="AAB96958.1"
/db_xref="GI:2791847"
/translation="MTILVVTGTGTGKTVVCAALASAAQAGIDVAVCKPVOVTGTA
RITDLLAEVRLAGVTOQLAGLARYPQPMAPAPAAAEHAGMALPAKDUIVRLIADLDRPG
RLTLLVAGAGLLVEIAEPGVTURDVAADVAAALVYVVTADICTLNNHILKLTLEALAAQ
VSCAGLVIGSWPDPFPCLVAAASNRSAIARIATVTRALPAGAASLIDAGITFAASAAAFDR
NNVAGLVG"
4957..5466
/note="ORF6: similar to MTCY336.32c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96959.1"
/db_xref="GI:2791848"
/translation="MVHSIELVFDSDTEAIIIRRIWAGLAAAGIPQAPASRPHVSLAV
AERIAPEVDEPLGAVARRLPDLCVIGAPVLFGRANVVTRLVVPTSEILALHAETHRL
CGHPLAPAPKANSIPQWTAHVTLARRVGGHQLGRALIRIAGRPSRIICRGFAGLRWDG
NTHAEYLLG"
complement(5612..6478)
/note="ORF7: similar to MTCY251.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96960.1"
/db_xref="GI:2791849"
/translation="MDWLHPDGLDTERARKRGITLSNQYDGMSELSCYLTPOARA
TFEAVILAKLAAPGATNPDDHTPVIDTTPDAAAIDRTRSQQRNHGHTIAGLRALLAS
GELGQHNGLPVSIIVTTTLDLTQGTAGKFTGGTLLPQADVIRMTSHHHYSPASGR
YQVAFIDHGTPIALYHTKRLASPAQRIMLFANDRGCTKPCCDAPAYIISQAHVHTGMS
TGHTDITDLTCLADPDNRLAEKWTTRKNTHTGHTWLPPLPHDHCQIRNTTFHHHEKL
LRHNEDNHJDP"
complement(6339..7007)
/note="ORF8: similar to MTCY336.16"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96961.1"
/db_xref="GI:2791850"
/translation="MIANSREELVEFDALDAELDRLDEVSFEVLTTPERLSLERLE
CIYVRLPAVGHTLINDLTQVASEELGGTLCALANLRITKPDAAIRIADADLGP
KAITGEPALQITATATAQOGLIGEAIKIVIRALFPKARRGCCVHIPGRRSRBQ
SRSTSSRRAGPLPAGHGLATPRRRPHRTRPQTRHHPEOPAIRRHVTAKWLPDPQS
AGHL"
7455..8504
/genes="bioH"
/genes="bioH"
/genes="bioH"
/note="BioH: similar to MTCY346.15c"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAB96962.1"
/db_xref="GI:2791851"
/translation="MTQAATPTNDAGQDGNNSDILLVVARQOVLQKCGINODQVLA
VIQLPDDRLELLALAHFVRMRWCQPEVEGIIISLKTGGCPELCHFCQSGLFPASPV
RSWALDIPSLVEAAKQAKSGATEFCIVAARGPDERLMAQVAAIGIETAKRNEVEINIA
CSIGMLTAEOVDQDLAARGVIIRYNNHNTSFANVYVTHTWEEKWTI.SMKVDGME
VCCGILTMGQFTLOOARAEFAELAEGLGDEVPFLNPRPCTPFADILEVMPVGIKAL
VAAERIALPRTMILKPAAGREITLGDLGAKKGLGGINAVIVGNVLTTLGNPAEADLEI
IDELQMLPKALNASLI"
8825..9406
/note="ORF10: similar to MTCY336.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96963.1"
/db_xref="GI:2791852"
/translation="MLGSATGVLVGLWMIAPPIHVAIVAITRAGERVHEYLGSSEQ
NFFIAPFMLGLLSVLYAVASALMQWREHGPQWVAGLSIGLITAAAIAGVAGLVV
RLRYGALDFDTVPLSRGSHALTYVTPAPVFFARRPQIALTLMPAGIASLVYALLA
ACTARDLGGYPADVDPSSNARTEALETPQAPVS"
complement(9571..10019)
/note="ORF11: similar to MTCY336.12"
/codon_start=3
/transl_table=11
```


KEYWORDS
SOURCE
ORGANISM

Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS

1 (bases 1 to 32437)
Cole, S.T., Brosch, R., Parkhill, J., Garner, T., Churcher, C.,
Mole, S.T., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaisa, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Konar, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and
Barrell, B.G.

TITLE

Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Nature. 393 (6685), 537-544 (1998)
98295987
2 (bases 1 to 32437)
Parkhill, J.

Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75224 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2117233.

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES
source

Location/Qualifiers
1. .32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1. .273

source

/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y48"
33. .32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y336"

source

complement(39. .2228)
/gene="Rv1565c"
complement(39. .2228)
/gene="Rv1565c"
/note="Rv1565c", len: 729. Unknown membrane
protein, some similarity to O05402 HYPOTHETICAL 72.2 KD
PROTEIN from B. subtilis (634 aa) opt: 384 E(): 4.8e-17;
29.1% identity in 378 aa overlap and to hypothetical
protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, Y392_HAFIN P43993 h10392 (245 aa) opt: 265;
E(): 5.5e-10; 28.3% identity in 247 aa overlap. TParse
score is 0.930"

gene

CDS

complement(39. .2228)
/gene="Rv1565c"
complement(39. .2228)
/gene="Rv1565c"

gene

CDS

complement(39. .2228)
/gene="Rv1565c"
complement(39. .2228)
/gene="Rv1565c"

/note="Rv1568, (MTCY336.35C), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC:2.6.1.62) O06622. Contains
aminotransferases class-III pyridoxal-phosphate attachment
site (P500600). FASTA results, bioA_MyCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
identity in 421 aa overlap. TParse score is 0.874."

/transl_table=11
/product="hypothetical protein Rv1565c"
/protein_id="CAB09072.1"
/db_xref="GI:2117272"
/db_xref="SPTREMBL:O06625"
/translation="MLTSPRRPALTPPALPVTMTGTRTTGFRHLDGURGVAlA
LVAVPHWFRGVSQGVDFLALSGFFGKILRAALNPDLSPISPIAEVIRLRLLPA
LVVLAGCALLTIAIQPTTWREAFANOSLASGYQNMELASTVSNYLRAGAVSPQI
HWSMSVQCOFYLAFLLLVAGGAYLLRLFRGRPRAPYLRTMFVLLSTITIASFYAI
VAHIAQATAYNTFAKELNELLAGALVGVVHVRWPHLRTAVATAALATLSCGAL
IDGKFFPGPMALVPVGTAMITLAGANRQGHFGTRDLPLNRLLEDVLRPRAPGV
YSYLWHFPLLI FWSYTGHRHANFVTEAALLVGLLAYLTLRLVEDVLRPRAPGV
RSPAAYVPPIWRLRRPTVLGVSVALTATSTFTRHREHIVQRAAGKELSGLS
SRDPCARALIDHVRVPLKMRPTVLELZDLPTSTKDCI SDFVNPALINCTYGDVD
APTIALAGSHAELHMLTALDILGRMHKVVYLYLKMGCLPISDFVPLJMGNNAPYQ
CHWQQAAMAKLVADHPDYVFTSTRPNKIPGDVMPATYVGIWOTFADNNTPLVLA
DTWLVKGOFFPADCLAKGNQSCGIARSLVLDNRPTLDFAKFFLLKPLDMSD
AICRTDTCRAVEGNVLVDRSHLTPYMTWTSELGKQIAANTDMW"
complement(2327. .3019)
/gene="Rv1566c"
complement(2327. .3019)
/gene="Rv1566c"
/note="Rv1566c, (MTCY336.37), len: 230. Function: unknown
probably exported has QQAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasins
1,2(O07390, O07391) from M. avium and slightly similar to
C-terminus of *Listeria* invasion-associated protein p60
precursor p60.LISMO P21171. FASTA results, O49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. TParse score is 0.890"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAH09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:O06624"
/translation="MKRSMKSGSFATGLAMWLAPMVAAPGLAAADPATRPVDVQQITD
VVIATLQSGVPSWAGGIGTGTGTGINTVGFDSGLLIQYAYAGAGLIPRSS
QMYKVQKVLPOQARKDLIFYPGEGTQVALYLGKQMLVEDGVVQVSPVPTNGMT
PYLVILGTQPTPVQAPVPAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAP
VQPPFGTARS"
complement(3030. .3036)
/note="possible RBS. AAGGAGG, for Rv1566c"
complement(3258. .3263)
/note="ASNI site: ATTAAT; probably linking fragments
H3/G36"
complement(3259. .3543)
/note="Rv1567c"
complement(3259. .3543)
/gene="Rv1567c"
/note="Rv1567c, (MTCY336.36), len: 94. Unknown. TParse
score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:O06623"
/translation="MVTMTSPSRPLFAFTDNVCPDAPLVPGVNYLYPVMWGGIG
AAIATAVIGPVSMLKGYMSFWPIISIAVITVTSIAGYATAGFSRWH"
3791. .5104
/gene="bioA"
3791. .5104
/gene="bioA"
/note="Rv1568, (MTCY336.35C), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC:2.6.1.62) O06622. Contains
aminotransferases class-III pyridoxal-phosphate attachment
site (P500600). FASTA results, bioA_MyCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
identity in 421 aa overlap. TParse score is 0.874."

[illegible]

274410.1 GI:3261600

. Mycobacterium tuberculosis H37Rv.
 Mycobacterium tuberculosis H37Rv
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 38380)
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
 Hamlin, N., Hawley, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 RaJandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellon, S.,
 Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrell, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature. 393 (6685), 537-544 (1998)
 98295987
 2 (bases 1 to 38380)
 Parkhill, J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique
 Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1405752.
 Notes:
 details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of 78 genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.
 Location/Qualifiers
 1..38380
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 <1..35889
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y251"
 6..956
 /gene="hycD"
 6..956
 /gene="hycD"
 /note="RV0084, (MTCY251.02), len: 316, hycD, similar to
 HYCD_ECOLI P16430 formate hydrogenlyase subunit 4 (307 aa)
 FASTA scores, opt: 570, z-score: 591.6, E(): 2.1e-26,
 (33.8% identity in 305 aa overlap) and NUOH_ECOLI
 p33603nadh dehydrogenase i chain h (325 aa) FASTA scores,
 opt: 207, z-score: 220.7, E(): 9.5e-06, (26.5% identity in
 260 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hycD"
 /protein_id="CAA98920.1"
 /db_xref="GI:1405754"
 /db_xref="SRRPWR1.010881"

```
/translation="MSYLAGAAQICGVWVCAPIVIGMTROVRAHWEGRAGAGLLOPHWR
DLIKOLGQQOITPAGCTITIVFAAPVIVAGITITLLLAETIAPLVATGSPIDPDSADLPVAVG
LIFTGTAITLGIADTGTSGFGMSASREITIAALVEPTITILLAFALSIAGSANGIAL
VASTIDHCVSLAGVLAVFALVIVIAETGRPLVPDNPATHELTIVHVEAMVLEYAG
PRALVHVEAGMRLTIVLLALLANLFLPWGIAAGAPTALDVLITGVVAARAAVAILAVLL
ATFEVFLAKRLFRVPEILLAGSFLALLALLAATAANFETVGA"
967..1629
/gene="hycp"
CDS
967..1629
/gene="hycp"
/notes="Rv0085, (MTCY251.03), len: 220. hycP, similar to
sp|P77524|HYFE_ECOLI HYDROGENASE-4 COMPONENT E (216
aa); FASTA scores: opt: 204 z-score: 272.1 E(): 1.2e-07;
25.5% identity in 216 aa overlap"
/codon_start=1
/transl_table=11
/product="hycp"
/protein_id="CAA98921.1"
/db_xref="GI:1405755"
/db_xref="SWISS-PROT:Q10882"
/translation="MSNANFSILVDFAGGLVLASVLIWVRDLRAIVRLAWOGAAL
AAPIILGRINDRALIAVGIALLRALVLPMLLARAVGAEEAAQREATPLVNTASS
LITAGLTAFATITQPVVNEPQVTINAVPAFAVLLIATFVMTIRLUHVAVSQAGFL
MLDNGIATAFLITAGVPLIVELGASLDVLFVAVIVIGVLTGRLRLIFGDADDLKREL
RP"
1629..3095
/gene="hycQ"
CDS
1629..3095
/gene="hycQ"
/notes="Rv0086, (MTCY251.04), oxidoreductase, len: 488,
hycQ, most similar to sp|P77437|HYFE_ECOLI HYDROGENASE-4
COMPONENT F(EE 1.- (526 aa), FASTA scores: opt: 948
z-score: 1117.4 E(): 0: 35.9% identity in 493 aa overlap.
Also similar to E. coli d9087711 & NUOL_ECOLI P33607 nadh
dehydrogenase chain 1 (613 aa) FASTA scores, opt: 360,
z-score: 354.9 E(): 3.2e-13, (27.9% identity in 488 aa
overlap), and to NUOL_ECOLI P33608, nadh dehydrogenase i
chain n (425 aa), FASTAScores, opt: 375, z-score: 371.4,
E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/db_xref="SPTREMBL:Q10883"
/translation="MTGLLILAILAPLAASLITGWRRTATITATLSATTVLACAV
AMGFWMGSGAQFGLGGLRADALTVMVLVIGVGTPLATAASIGYIDTELAIGHIUGR
SARLYGLTPAFICAMVLAVCANNICVIVIAETATITAFIVGHKRTALEATWKY
VVICSVGIAPFELGTVLVLYFAARDGAAAAAGALNDILAEHAAGLIDPGVARLUGGLLL
IGYKAGAGLPFPHTLADAHQAPAPVSAIMSGVLLAVAFSVLIHLRPLIDAVSGPAY
LRNGLLVGLATLVALLVMTVTGDKPKMLAYSMEHMGIIATAAAGTTLAIAALLL
HLVHAGIGKTVLFLAGSOLQAADHSTAIADITGVMRSRSLIGVSFAVGLIVLLGPPF
MTAFSELATARSLEANERLANWLGAAALLIITAGFTALARNRSMRLTGTTPAAGPATIVP
ATAAALMWGIVVSAALGITAGTADLLGLLGAASNVGLP"
3092..4570
/gene="hycE"
CDS
3092..4570
/gene="hycE"
/notes="Rv0087, (MTCY251.05), oxidoreductase, len:
492,hycE, similar to HYCE_ECOLI P16431 formate
hydrogenlyase subunit 5 (569 aa) FASTA scores, opt: 680,
z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa
overlap) and to NUOL_ECOLI P33600 nadh dehydrogenase i
chain d, (407 aa) FASTA scores, opt: 245, z-score: 293.1,
E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/db_xref="SPTREMBL:Q10884"
/translation="MMSASWLRHRSVERGLIATAEOLWADSFRLAIVAAHDDGDSLRV
VYLFAGYPPRRVELEYVVPADNPERSLAYLSFPAGRERENADYILIRPVGHKPRR
```

```
RLVRHAHPDWHPMRTDAGPAPEFTTDTGAPPELAVEGPGVYEIIPVGVHAGLIEPGHF
RESVAGETIVRLKARLWVHRGTEKLFPHGRPATAVDAVDLAERISGDT:SAAHALSHSLAI
EDALGIELPHEVHRLALIVELERLYNHAADGALANDGYSLANAIAOKIRENLRLR
NAAVTGHRLRGALIRAGVALRALPDOTDELAALAVLATVATLANSVYDRFACTA
VLIPDDASALGCGIYVARASGLRSDRDELAAVLITELGAPGIVLAKYIVVRDEF
AASAAIAQHIHIVESHTGP IEYAATLHPVGAPSSGIGIVEGMRGTI VHKVF:IDVDGRITR
AKVVDSPMFNNPAILPVAMADTIVPDFPLANKSFNQSYAGNDL"
4605..5279
/gene="Rv0088"
CDS
4605..5279
/gene="Rv0088"
/notes="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/db_xref="SWISS-PROT:Q10885"
/translation="MSYKHPASRVLRQTRSTVVKRSGSLSNRRVYRTGDLGLAVWG
GREYRAVKPCTPGIQPKGDMVTIVVDAGPGRVSRSEVAAPAAEI:FAIVADPRRHR
ELDGSQTVRGNIKVPARKLVVGSKFSTKMLFGLPYRITSRTALKPNELVWESHPLGII
RWWEFESLSPTLTRVTETEDYHAAGAIAKNGLKFYEMTGFAKSNAAI:FEATLAKLSQ
YAKGRA"
5436..6029
/gene="Rv0089"
CDS
5436..6029
/gene="Rv0089"
/notes="Rv0089, (MTCY251.07), len: 197. unknown, some
similarity to sp|P12999|BIOC_ECOLI BIOTIN SYNTHESIS
PROTEIN BIOC. (251 aa). FASTA scores: opt: 202 z-score:
262.0 E(): 4.5e-07; 39.0% identity in 118 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0089"
/protein_id="CAA98925.1"
/db_xref="GI:1405759"
/db_xref="SWISS-PROT:Q10886"
/translation="MDQPWANIHYDALLDAMVPLGTQCVLIDVCCGIGLLAARLARRI
PYTAVDIDAPVILKRAQTRFANAP:IRMLHADIMTAELPNAGFDVAV:NAALHHIEDTR
TALSRLGLVTPGGT LAVVTEVTPSLRNLGHLSTSWACGMANRVKKGWHSAPIKWP
PPQTLHELURSHVRALLPGACIRRLLYGRVLVTRAPV"
6147..6152
RBS
6158..6928
/gene="Rv0090"
CDS
6158..6928
/gene="Rv0090"
/notes="Rv0090, (MTCY251.08), len: 256. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0090"
Query Match 38.9%; Score 176; DB 1; Length 38380;
Best Local Similarity 100.0%; Pred. No. 2.2e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 278 cgcggcccaaccttgaagccgtgctagcacaactgcccgcggcgagcaaacccq 337
Db 12127 CGCGGGCCACTTTGAAGCGGTGCTAGCCAAACTGCCGCCGCCGACCAACCCG 12068
QY 338 acgaccacaccccgctcatcgacacaccccgatcgcccgccatcgccgcgaccc 397
Db 12067 ACGACCACACCCCGGTATCGACACCAACCCCGATCGCGCGGCATCGACCGGACACC 12008
QY 398 gcgccaagcccccaacacacacgacgggctgctgcccgggctgcgcgcgtgac 453
Db 12007 GCAGCCAAAGCCCAACGCAACACGACGGGCTGCTGCGCGGGCTGCGCGGCTGATC 11952
RESULT 14
MTY13E12 43401 bp DNA BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
DEFINITION
```

ACCESSION	295390	AL123456	
VERSION	295390.1	GI:3261766	
KEYWORDS			
SOURCE	Mycobacterium tuberculosis H37Rv.		
ORGANISM	Mycobacterium tuberculosis H37Rv		
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
AUTHORS	1 (bases 1 to 43401) Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekle, A., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellion, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.		
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
JOURNAL	Nature, 393 (6685), 537-544 (1998)		
MEDLINE	98295987		
REFERENCE	2 (bases 1 to 43401)		
AUTHORS	Parkhill, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk		
COMMENT	On Jun 27, 1998 this sequence version replaced gi:2104370. Notes: Details of M.tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of 79 genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-11bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.		
FEATURES	source	Location/Qualifiers	
		1..43401	/organism="Mycobacterium tuberculosis H37Rv"
		/strain="H37Rv"	
		/db_xref="taxon:83332"	
	source	<1..1481	/organism="Mycobacterium tuberculosis H37Rv"
		/strain="H37Rv"	
		/db_xref="taxon:83332"	
		/clone="Y77"	
	gene	78..1481	/gene="Rv3448"
		78..164	
	misc_feature	/gene="Rv3448"	
		/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"	
		78..1481	
	CDS	/note="Rv3448"	
		/gene="Rv3448"	
		/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature. Some transport systems inner membrane comp signature. Similarity to AL021930 MTV035.18 Mycobacterium tuberculosis (472 aa) fasta score, opt: 429 z-score: 488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein Rv3450c"	
		/product="hypothetical protein Rv3448"	
		/protein_id="CAB08736.1"	
		/db_xref="GI:3261767"	
		/db_xref="SPTREMBL:O33354"	
		/translation="MPTSDPGLRRVTVTVHAGAAQAVDLTLPAAVFPVATILPSIVDILGDR CASPATAARYQVLSALCAPALPNATTLAOCGIRDGAVILVLRKSSAQHPPTCCDDVAFAV AAALDTPARQQCKTTLSCALAAACITAGGGLMLVRLNLTGNVTKYSUATAGVVAFAA GIALDFAVITACRTYRDPIAGLTSLVIATIFGAVAGLLAIVPGVKGVHIVLVAAMAAA TSVLAMRITCGGGLTITAVACCACAVVAARATLVGAITAAPVPAIGSLATIASGLLEVS ARKAVLLAGLSPIRLPALNPADADALPTLDRITRANRADAMITSLIAAFAASATFVA IGTAVATHGIIHRSMSGGIALAAVTGALLLLARSADYTRSLVFAICGTTTATATFVA ADNALEHPWIAATITAMLAAMAVFIQFVAPALSLSPTVYRTTLLHETLALIAWVPLTA WICGAYSVAVRHLDLTWT"	
		611..>43401	
	source	/organism="Mycobacterium tuberculosis H37Rv"	
		/strain="H37Rv"	
		/db_xref="taxon:83332"	
		/clone="Y13E12"	
	gene	1478..2845	
		/gene="Rv3449"	
	CDS	1478..2845	
		/gene="Rv3449"	
		/note="Rv3449, (MTCY13E12.02), len: 455. Function: probable precursor of serine protease. Has putative signal peptide N-terminus and hydrophobic stretch at C-terminus. Contains three signatures typical of subtilase family: aspartic acid active site (PS00136), histidine active site (PS00137), serine active site (PS00138). FASTA results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E(): 8e-07; (38.0% identity in 387 aa overlap), similar to MTCY15F10.29 (45.7% identity in 451 aa overlap)"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein Rv3449"	
		/protein_id="CAB08735.1"	
		/db_xref="GI:2104372"	
		/db_xref="SPTREMBL:O06316"	
		/translation="MTSKTLRLVVSALATLSGLGTVAHAVSPPIIDERMPLPESAL PRRPTVQREVCTEVTAEGRAGRSQAQLADQVWRLTRGASQRAVIDTGVA RHRLPLKPVAGDGYFTGTGATDCDAHGTLVAGLIAAPDASDNFVCAAPADPDRAL RUSSKFAFVGPDPSTGVDVDMAKAVRTADLGCASVINISSLVAPADYVLTGCS GAALAYAVIUKNAVIAAAGNTGGAQCPOAPGVTRDSVTVAVSIAWYDDYVLTGCS VNAQKHPSAFTLAFHVIDVAATGCAVTSLSPPGCTGVNRKLRQHSIPISOTSAAPV VSGLAALHARTTLLTAKVMORIESTAHHPACWDPLVGNCTVIALAAVSSDSIPQA GLNLSIPALVAVIVLRKSLHGPSHKKALHTATPAGAACTLLAIPATLATASRRIRPGRN GIALGP"	
		1/57..1789	
	misc_feature	/gene="Rv3449"	
		/note="PS00136 Serine proteases, subtilase family, aspartic acid active site"	
		1862..1894	
	misc_feature	/gene="Rv3449"	
		/note="PS00137 Serine proteases, subtilase family, histidine active site"	
		2456..2488	
	misc_feature	/gene="Rv3449"	
		/note="PS00138 Serine proteases, subtilase family, serine active site"	
		complement(2810..4222)	
	gene	/gene="Rv3450c"	
		complement(2810..4222)	
	CDS	/gene="Rv3450c"	
		/note="Rv3450c, (MTCY13E12.03c), len: 470. Function: unknown but similar to Rv3895c MTCY15F10.17 (FASTA score: 1.6e-27; 36.4% identity in 475 aa overlap): possible membrane spanning region near N-terminus, and to Y14967 MLCB28_16 Mycobacterium leprae cosmid H628; (481 aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 32.9% identity in 480 aa overlap, also similar to Rv3869 (MTV027.04)"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein Rv3450c"	

[illegible]

us-09-785-904-2.olig.rge

Wed Apr 3 08:33:13 2002

Job time: 11239 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: April 3, 2002, 01:08:11 ; Search time 172.6 Seconds
(without alignments)
2250.106 Million cell updates/sec

Title: us-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccgggctgcgcgctatc 453

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				N_Geneseq_1101.*	
1:	/SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*	2:	/SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*	3:	/SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*	5:	/SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*	6:	/SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*	8:	/SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*	9:	/SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*	11:	/SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*	12:	/SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*	14:	/SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*	15:	/SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*	17:	/SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*	18:	/SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*	20:	/SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*	21:	/SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*				

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	176	38.9	650	17	M. Tuberculosis ma
2	176	38.9	650	19	DNA for M. tubercu
3	176	38.9	1535	16	Mycobacterium tube
4	176	38.9	1535	19	M. tuberculosis ce
5	176	38.9	1535	19	DNA for M. tubercu
6	176	38.9	12412	17	BCG deletion regio
7	18	4.0	1413	22	C glutamicum codin
8	18	4.0	1413	22	Corynebacterium gl
9	18	4.0	1536	21	Bordetella pertuss
10	18	4.0	1853	6	Plasmid pTnp1-35 b
11	18	4.0	1853	10	Sequence including

12	18	4.0	8051	18	AAT72685	Sugar biosynthesis
13	18	4.0	8160	20	AAAX5774	S.crythraea erythr
14	18	4.0	35026	21	AAA64890	Bordetella pertuss
15	18	4.0	34980	22	AAH68526	C glutamicum codin
16	17	3.8	281	21	AAH98508	Human colon cancer
17	17	3.8	300	20	AAAX98072	Nucleotide sequenc
18	17	3.8	306	20	AAAX98073	Nucleotide sequenc
19	17	3.8	441	22	AAH10473	Probe #406 for gen
20	17	3.8	441	22	AAH131455	Probe #413 used to
21	17	3.8	441	22	AAH131727	Probe #413 used to
22	17	3.8	555	22	AAH14923	Probe #1609 used
23	17	3.8	588	20	AAAX98044	Nucleotide sequenc
24	17	3.8	742	13	AAQ20926	C10-E15 DNA fragme
25	17	3.8	932	13	AAQ20923	C10-E12 DNA fragme
26	17	3.8	957	20	AAAX98046	Nucleotide sequenc
27	17	3.8	1077	22	AAH60764	Pseudomonas sp ABC
28	17	3.8	1188	22	AAH67681	C glutamicum codin
29	17	3.8	1410	17	AAH09251	SKK phosphotyrosin
30	17	3.8	1605	20	AAAX98068	Nucleotide sequenc
31	17	3.8	2172	15	AAQ56930	Bacillus thuringie
32	17	3.8	2372	22	AAH59617	Human polynucleoti
33	17	3.8	2444	22	AAH60835	Pseudomonas sp exp
34	17	3.8	2592	22	AAH26409	Pseudomonas sp hea
35	17	3.8	2634	19	AAH22334	Microbisporea therm
36	17	3.8	3048	22	AAH65828	C glutamicum codin
37	17	3.8	3141	20	AAAX98041	Nucleotide sequenc
38	17	3.8	4176	22	AAH57831	Human polynucleoti
39	17	3.8	4176	22	AAH57831	Human polynucleoti
40	17	3.8	5054	22	AAH07024	Pseudomonas putida
41	17	3.8	5544	22	AAH24891	Pimaricin biosynth
42	17	3.8	16382	16	AAH63074	IgG-FC binding pro
43	17	3.8	42235	20	AAAX98035	Nucleotide sequenc
44	17	3.8	42235	20	AAAX98035	Nucleotide sequenc
45	17	3.8	42235	20	AAAX98035	Nucleotide sequenc

ALIGNMENTS

RESULT 1	
AAH33658	
ID	AAH33658 standard; DNA; 650 bp.
AC	AAH33658;
XX	
XX	22 MAY-1996 (first entry)
XX	
DE	M. tuberculosis macrophage survival gene.
XX	
XX	Cellular uptake; cell entry; macrophage survival; vaccine;
KW	passive immunisation; gene therapy; ds.
XX	
OS	Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
XX	
PN	W09626275-A1.
XX	
PD	29-AUG-1996.
XX	
PF	20-FEB-1996; 96W0-US02155.
XX	
PR	22-FEB-1995; 95US-0392210.
XX	
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Riley LW;
XX	
DR	WPI; 1996-425086/42.
DR	p-PDB; AAH02303.
XX	
PT	DNA giving M. tuberculosis ability to enter mammalian cells
PT	and/or survive within macrophages; useful in vaccines to protect
PT	mammals against Mycobacterium tuberculosis infection
XX	

PS Claim 7; Page 45; 67pp; English.

XX A DNA molecule (AAT33658) codes for a protein (AAW02303) that confers

CC on Mycobacterium tuberculosis an ability to survive within

CC macrophages. It is a fragment of a larger molecule (AAT33656) obtd.

CC by ligating M. tuberculosis genomic DNA fragments into pBluescript

CC II vector and screening recombinant E. coli strains for HeLa cell-

CC invasive clones. It can be incorporated into a vector and used

CC for prodn. of recombinant macrophage survival protein, which is

CC useful in vaccines or for facilitating uptake of other materials,

CC e.g. therapeutic genes, into mammalian cells. A second DNA molecule

CC (AAT33657) codes for a protein (AAW02302) that confers on M.

XX tuberculosis the ability to enter mammalian cells.

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 38.9%; Score 176; DB 17; Length 650;

Best Local Similarity 100.0%; Pred. No. 4.6e-77;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggacacctttgaacgctgtagccaaactggcccccgcgcgacccaaccccg 337

Db 221 cgcgggacacctttgaacgctgtagccaaactggcccccgcgcgacccaaccccg 280

Qy 338 acgaccacaccccggtcatcgacaccccccgcgatgcgcccgcacatcgacgcgacaccc 397

Db 281 acgaccacaccccggtcatcgacaccccccgcgatgcgcccgcacatcgacgcgacaccc 340

Qy 398 gcagccaagcccaacgcaacacacaccccggtgctggccgggctgcgcgctgctgac 453

Db 341 gcagccaagcccaacgcaacacacaccccggtgctggccgggctgcgcgctgctgac 396

RESULT 2

AAV18649

ID AAV18649 standard; DNA; 650 BP.

XX

AC AAV18649;

XX

DT 03-JUL-1998 (first entry)

XX

DE DNA for M. tuberculosis cellular uptake protein fragment.

XX

KW Cellular uptake protein; vaccine; infection; ds.

XX

OS Mycobacterium tuberculosis.

XX

FH Key Location/Qualifiers

FT CDS 1..649

FT /*tag= a

FT /note= "stop codon not given"

XX

PN W09805784-A1.

XX

PD 12-FEB-1998.

XX

PF 06-AUG-1997; 97WO-US13056.

XX

PR 07-AUG-1996; 96US-0689411.

XX

PA (CONN-) CONNAUGHT LAB LTD.

PA (CORR) CORNELL RES FOUND INC.

XX

PI Chong P, Riley LW;

XX

DR WPI; 1998-145620/13.

DR P-PSDB; AAW47543.

XX

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of

PT therapeutic agents e.g. antibiotics, also useful in vaccines

XX

PS Disclosure; Page 14; 82pp; English.

XX

XX The present sequence encodes a Mycobacterium tuberculosis cellular

CC uptake protein fragment, which confers on M. tuberculosis an

CC ability to survive within macrophages.

CC The protein can be used in a vaccine to prevent M. tuberculosis

CC infection, and provide for the uptake in cells of, e.g.

CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies

CC raised against it can be used to treat mammals already exposed to

CC M. tuberculosis, to induce a passive immunity and prevent disease

CC occurrence.

XX

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 38.9%; Score 176; DB 19; Length 650;

Best Local Similarity 100.0%; Pred. No. 4.6e-77;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggacacctttgaacgctgtagccaaactggcccccgcgcgacccaaccccg 337

Db 221 cgcgggacacctttgaacgctgtagccaaactggcccccgcgcgacccaaccccg 280

Qy 338 acgaccacaccccggtcatcgacaccccccgcgatgcgcccgcacatcgacgcgacaccc 397

Db 281 acgaccacaccccggtcatcgacaccccccgcgatgcgcccgcacatcgacgcgacaccc 340

Qy 398 gcagccaagcccaacgcaacacacaccccggtgctggccgggctgcgcgctgctgac 453

Db 341 gcagccaagcccaacgcaacacacaccccggtgctggccgggctgcgcgctgctgac 396

RESULT 3

AAQ89200

ID AAQ89200 standard; DNA; 1535 BP.

XX

AC AAQ89200;

XX

DT 09-OCT-1995 (first entry)

XX

DE Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry

DE protein.

XX

KW Vaccine; tuberculosis; ss.

XX

OS Mycobacterium tuberculosis.

XX

FH Key Location/Qualifiers

FT CDS 1..1535

FT /*tag= a

XX

PN W09506726-A.

XX

PD 09-MAR-1995.

XX

PF 01-SEP-1994; 94WO-US09863.

XX

PR 02-SEP-1993; 93US-0118442.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Riley LW;

XX

DR WPI; 1995-115442/15.

DR P-PSDB; AAR71931.

XX

XX DNA encoding for cellular uptake of Mycobacterium tuberculosis

PT used to develop prods for vaccines, passive immunisation and

PT diagnosis and cellular uptake of other materials

XX

PS Claim 2; Page 9-11; 46pp; English.

XX

CC The isolated DNA molecule of the invention confers on M.

CC tuberculosis an ability to enter cells and to survive within

CC

CC	A DNA molecule (AAAT33656) confers on Mycobacterium tuberculosis an ability to enter mammalian cells and to survive within macrophages.									
CC	The encoded protein sequence is given in AAW02301. The DNA was obtd.									
CC	by ligating M. tuberculosis genomic DNA fragments into pBluescript II									
CC	vector and screening recombinant E. coli strains for HeLa cell-									
CC	invasive clones. The DNA includes 2 separate coding regions (see									
CC	also AAAT3657-58) coding for the cell entry (AAW02302) and macrophage									
CC	survival (AAW02303) proteins. It can be used to produce the cellular									
CC	uptake proteins used as vaccines or to facilitate uptake of other									
CC	materials, e.g. therapeutic genes.									
XX	Sequence 1535 BP; 247 A; 544 C; 458 G; 236 T; 0 other;									
SS										
Query Match	38.9%; Score 176; DB 17; Length 1535;									
Best Local Similarity	100.0%; Pred. No. 4.3e-77;									
Matches 176; Conservative	0; Mismatches 0; Indels 0; Gaps									
QY	278	cgcggccaccctt	tgaagcctgtctagc	caactggccgccccgcgcgacacccgc	337					
Db	1106	cgcggccaccctt	tgaagcctgtctagc	caactggccgccccgcgcgacacccgc	1165					
QY	338	acgaccacaccc	ggtcatcgacacacaccccgatgcgcgcgcacatgcgcgcgcacac	397						
Db	1166	acgaccacaccc	ggtcatcgacacacaccccgatgcgcgcgcacatgcgcgcgcacac	1225						
QY	398	gcagccaaagcc	caaccgcgaaccacgcggctgctgcgcgggctgcgcgcgcgcgcgc	453						
Db	1226	gcagccaaagcc	caaccgcgaaccacgcggctgctgcgcgggctgcgcgcgcgcgcgc	1281						
RESULT	5									
AAV18647										
ID	AAV18647 standard; DNA; 1535 BP.									
XX										
AC	AAV18647;									
XX										
DT	03-JUL-1998 (first entry)									
XX										
DE	DNA for M. tuberculosis cellular uptake protein fragment.									
XX										
KW	Cellular uptake protein; vaccine; infection; ds.									
XX										
OS	Mycobacterium tuberculosis.									
XX										
FH	Location/Qualifiers									
FT	1..1534									
FT	/*key a									
FT	/note= "Stop codon not given"									
XX										
PN	W09805784-A1.									
XX										
PD	12-FEB-1998.									
XX										
PF	06-AUG-1997; 97W0-US13056.									
XX										
PR	07-AUG-1996; 96US-0689411.									
XX										
PA	(CONN-) CONNAUGHT LAB LTD.									
PA	(CORR) CORNELL RES FOUND INC.									
XX										
PI	Chong P, Riley LW;									
XX										
DR	WPI; 1998-145620/13.									
DR	P-PSDB; AAW47541.									
XX										
PT	Mycobacterium tuberculosis DNA - confers ability to enhance uptake of									
PT	therapeutic agents e.g. antibiotics, also useful in vaccines									
XX										
PS	Disclosure; Pages 9-10; 82pp; English.									
XX										
CC	The present sequence encodes a Mycobacterium tuberculosis cellular									
CC	uptake protein fragment, which confers on M. tuberculosis an									

CC ability to enter mammalian cells and to survive within macrophages.
 CC The protein can be used in a vaccine to prevent M. tuberculosis
 CC infection, and provide for the uptake in cells of, e.g.
 CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
 CC raised against it can be used to treat mammals already exposed to
 CC M. tuberculosis, to induce a passive immunity and prevent disease
 CC occurrence.
 XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;
 SQ

Query Match 38.9%; Score 176; DB 19; Length 1535;
 Best Local Similarity 100.0%; Pred. No. 4.3e-77;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggacacattgaagccgtctagcaaaactggccgcccgcgcgacacaccccg 337
 Db 1106 cgcgggacacattgaagccgtctagcaaaactggccgcccgcgcgacacaccccg 1165

Qy 338 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacacaccc 397
 Db 1166 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacacaccc 1225

Qy 398 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacacaccc 453
 Db 1226 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacacaccc 1281

RESULT 6
 AAT33537 standard; DNA: 12412 BP.
 ID AAT33537
 XX AC AAT33537;
 XX DT 15-FEB-1998 (first entry)
 XX DE BCG deletion region 3 and flanking sequences.
 XX KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;
 XX KW mycobacteria; vaccine; infection; marker; ss.
 XX OS Mycobacterium bovis strain BCG.
 XX FH Key Location/Qualifiers
 FT misc_feature 1406..10673
 FT /*tag= a
 FT /note= "BCG delta 1 deletion region"
 XX PN WO9625519-A1.
 XX PD 22-AUG-1996.
 XX PF 15-FEB-1996; 96WO-US01938.
 XX PR 17-FEB-1995; 95US-0390878.
 XX PA (PATH-) PATHOGENESIS CORP.
 XX PI Mahairas GG, Stover CK;
 XX WPI; 1996-393419/39.
 XX Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 PT bacterial infection
 XX Example 1; Fig 3; 66pp; English.
 XX This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta3. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT3353 and AAT33536) have also been detected.
 CC Identification involved screening a BCG cosmid library with a

CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.
 XX Sequence 12412 BP; 2146 A; 427 C; 3876 G; 1917 T; 200 other;
 SQ

Query Match 38.9%; Score 176; DB 17; Length 12412;
 Best Local Similarity 100.0%; Pred. No. 3.7e-77;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggacacattgaagccgtctagcaaaactggccgcccgcgcgacacaccccg 337
 Db 737 cgcgggacacattgaagccgtctagcaaaactggccgcccgcgcgacacaccccg 796

Qy 338 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacacaccc 397
 Db 797 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacacaccc 856

Qy 398 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacacaccc 454
 Db 857 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacacaccc 912

RESULT 7
 AAH65816/c
 ID AAH65816 standard; DNA: 1413 BP.
 XX AC AAH65816;
 XX DT 26-SEP-2001 (first entry)
 XX DE C glutamicum coding sequence fragment SEQ ID NO: 851.
 XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX KW organic acid synthesis; ds.
 XX OS Corynebacterium glutamicum.
 XX FH P1108790 A2
 XX PD 20-JUN-2001.
 XX PF 18-DEC-2000; 2000EP-0127688.
 XX PR 16-DEC-1999; 99JP-0377484.
 XX PR 07-APR-2000; 2000JP-0159162.
 XX PR 03-AUG-2000; 2000JP-0280988.
 XX PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Taleishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI; 2001-376931/40.
 XX P-PSDB; AAG90597.
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX Claim 8; SEQ ID NO: 851; 246pp + Sequence Listing; English.
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

Wed Apr 3 08:33:14 2002

are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 1413 BP; 259 A; 350 C; 395 G; 409 T; 0 other;

Query Match 4.0%; Score 18; DB 22; Length 1413;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 214 tcacctgagcaaccagc 231
|||||
Db 1007 TCACCTGAGCAACCAGC 990

RESULT 8
AAAF71958/c
ID AAF71958 standard; DNA; 1413 BP.

XX AAF71958;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:411.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

XX WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

XX 01-JUL-1999; 99DE-1030476.

XX 02-JUL-1999; 99US-0142101.

XX 08-JUL-1999; 99DE-1031415.

XX 08-JUL-1999; 99DE-1031418.

XX 08-JUL-1999; 99DE-1031419.

XX 08-JUL-1999; 99DE-1031420.

XX 08-JUL-1999; 99DE-1031424.

XX 08-JUL-1999; 99DE-1031428.

XX 08-JUL-1999; 99DE-1031434.

XX 08-JUL-1999; 99DE-1031435.

XX 08-JUL-1999; 99DE-1031443.

XX 08-JUL-1999; 99DE-1031453.

XX 08-JUL-1999; 99DE-1031457.

XX 08-JUL-1999; 99DE-1031465.

XX 08-JUL-1999; 99DE-1031478.

XX 08-JUL-1999; 99DE-1031510.

XX 08-JUL-1999; 99DE-1031541.

XX 08-JUL-1999; 99DE-1031573.

XX 08-JUL-1999; 99DE-1031592.

XX 08-JUL-1999; 99DE-1031632.

XX 08-JUL-1999; 99DE-1031634.

XX 08-JUL-1999; 99DE-1031636.

XX 09-JUL-1999; 99DE-1032125.

XX 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 12-AUG-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX (BADI) BASF AG.

XX Pompejus M, Kroeber B, Schroeder H, Zelder O, Haberhauer G;

XX WPI: 2001-137957/14.

XX P-PSDB; AAB79839.

XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases.

XX Claim 3; Page: 759 /61; 1747pp; English.

XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.

XX Sequence 1413 BP; 264 A; 349 C; 394 G; 406 T; 0 other;

Query Match 4.0%; Score 18; DB 22; Length 1413;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 214 tcacctgagcaaccagc 231

|||||

Db 984 TCACCTGAGCAACCAGC 967

RESULT 9

AAA64886

ID AAA64886 standard; DNA; 1536 BP.

XX

AC AAA64886;

XX	02-FEB-2001	(first entry)
XX	Bordetella pertussis	coding sequence # 2.
XX	Type III secretion system;	virulence factor; pathogenicity island; ss.
XX	Bordetella pertussis.	
XX	Key	Location/Qualifiers
XX	CDS	1..1536
FT	FT	/*tag= a
FT	FT	/product= "Protein # 2"
XX	WO200037493-A2.	
XX	29-JUN-2000.	
XX	21-DEC-1999;	99WO-EP10297.
XX	21-DEC-1998;	98GB-0028217.
XX	(ULBR) UNIV LIBRE BRUXELLES.	
XX	Bollen A, Fauconnier A, Godfroid E;	
XX	WP': 2000-452178/39.	
XX	P-PSUB; AAB14148.	
XX	Novel polypeptides derived from Bordetella pertussis, useful for	
PT	treating and diagnosing Bordetella infection -	
XX	Example 2; Pages 154-156*	165pp; English.
XX	Bordetella pertussis possesses a type III secretion system. Type III	
CC	secretion systems allow bacteria to target virulence factors directly at	
CC	host cells. The present sequence is a coding sequence of B.	
CC	pertussis. No name has been specified for the present sequence. The	
CC	present sequence may be a housekeeping gene. A pathogenicity	
CC	island is a compact, distinct genetic unit carrying virulence genes. The	
CC	present sequence is located within a pathogenicity island (see AAA64890)	
CC	CC which also carries, a number of genes encoding proteins involved	
CC	in the type III secretion system of B. pertussis. See AAA64849-A64884	
CC	and AAB14111-B14146 for details of the coding sequences and proteins	
XX	identified in the pathogenicity island, of the present invention.	
XX	Sequence 1536 BP; 248 A; 546 C; 560 G; 182 T; 0 other;	
XX	Query Match	4.0%; Score 18; DB 21; Length 1536;
XX	Best Local Similarity	100.0%; Pred. No. 17;
XX	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps	
Qy	170	ccgacggcgacctcaccg 187
Db	593	ccgacggcgacctcaccg 610
RESULT 10		
AAN50261		
ID	AAN50261	standard; DNA; 1853 BP.
XX	AC	
XX	AAN50261;	
XX	14-JAN-1992	(first entry)
XX	Plasmid pTrp1-35	DNA encoding the 2,5-diketogluconate-reductase gene.
XX	2,5-Diketogluconate-reductase;	ascorbic acid; ss.
XX	Corynebacterium.	
XX	Key	Location/Qualifiers

FT	promoter	229..234	
FT		/*tag= a	
FT	promoter	/label= E. coli trp promoter	
FT		252..257	
FT		/*tag= b	
FT	RBS	/label= E. coli trp promoter	
FT		279..283	
FT		/*tag= c	
FT	CDS	296..3130	
FT		/*tag= d	
FT		/label= 2,5-DKG-reductase gene	
XX			
PN	EPI32308-A.		
XX			
PD	30-JAN-1985.		
XX			
PF	25-JUN-1984;	84EP-0304277.	
XX			
PR	14-JUN-1984;	84US-0620585.	
PR	28-JUN-1983;	83US-0508409.	
PR	28-JUN-1983;	83US-0508410.	
PR	28-JUN-1983;	83US-0508628.	
PR	14-JUN-1984;	84US-0620651.	
PR	14-JUN-1984;	84US-0620652.	
XX			
PA	(GETH) GENENTECH INC.		
XX			
PI	Estell DA, Light DR, Rasteter WH, Lazarus RA, Miller JV;		
XX			
DR	WPI; 1985-026545/US.		
DR	P-PSDB; AAP50247.		
PT	New pure 2,5 di-keto-gluconic acid reductase - prepd. by		
PT	recombinant DNA methods for ascorbic acid prodn.		
XX			
PS	Disclosure; Fig 4; 44pp; English.		
XX			
CC	The 2,5-DKG-reductase gene is prepared by recombinant DNA techniques.		
CC	2,5-DKG-reductase is used to convert 2,5-DKG stereoselectively into		
CC	2-KLG, which is an intermediate in the production of ascorbic acid		
CC	(vitamin C).		
XX			
SO	Sequence 1853 BP; 327 A; 645 C; 587 G; 294 T; 0 other;		
Query Match			
Best local similarity: 4.0%; Score 18; Dk 6; length 1854;			
Matches 18; conservative 0; mismatches 0; indels 0; gaps			
Qy	1/7	cgagctcagcagacacga 194	
Db	1039	cgagctcagcagacacga 1056	
RESULT 11			
AA	AA90631		
ID	AA90631 standard; DNA; 1853 BP.		
XX			
AC	AA90631;		
XX			
DT	22-JUN-1990 (first entry)		
XX			
DE	Sequence including 2,5-diketogluconic acid reductase gene and pTrp135		
DE	control regions.		
XX			
KW	2,5-diketogluconic acid reductase; 2,5-DKG reductase;		
KW	2-keto-L-gluconic acid; 2-KLG; ascorbic acid; vitamin C; ss.		
XX			
OS	Corynebacterium sp. ATCC 31090.		
XX			
key	Location/Qualifiers		
FT	promoter	229..234	
FT		/*tag= a	

```

FT      promoter      /label=E. coli Trp promoter
FT      252..257
FT      /*tag= b
FT      /label=E.coli Trp promoter.
FT      RBS           279..282
FT      /*tag= c
FT      mat_peptide   296..1129
FT      /*tag= d
FT      /product=2,5-DKG reductase
FT      XX
XX      EP305608-A.
XX      XX
XX      08-MAR-1989.
XX      XX
XX      01-JAN-1987; 87EP-0202624.
XX      XX
XX      28-JUN-1983; 83US-0508628.
XX      PR
XX      28-JUN-1983; 83US-0508409.
XX      PR
XX      28-JUN-1983; 83US-0508410.
XX      PR
XX      14-JUN-1984; 84US-0620651.
XX      PR
XX      14-JUN-1984; 84US-0620652.
XX      PR
XX      14-JUN-1984; 84US-0620585.
XX      XX
XX      (GETH ) GENENTECH INC.
XX      XX
XX      Light DR, Estell DA, Lazarus RA, Rastetter WH, Miller JV;
XX      PI
XX      WPI; 1989-069875/10.
XX      DR
XX      P-PSDB; AAP94624.
XX      XX
XX      Recombinant prodn. of 2,5-diketogluconic acid reductase - useful for
XX      PT converting 2,5-DKG stereoselectively into 2-ketogluconic acid, a
XX      PT precursor of vitamin C.
XX      XX
XX      Disclosure; fig 4; 2lpp; English.
XX      PS
XX      The enzyme encoded by the sequence, 2-KLG reductase is used to convert
XX      CC 2,5-DKG into 2-KLG, a precursor of ascorbic acid.
XX      CC See also AAN94347 and AAN94348.
XX      CC
XX      Sequence 1853 BP; 327 A; 644 C; 588 G; 294 T; 0 other;
XX      SQ

Query Match      4.0%; Score 18; DB 10; Length 1853;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 cgacctcaccgacaccga 194
Db 1039 cgacctcaccgacaccga 1056

RESULT 12
AAT72685
ID AAT72685 standard; DNA; 8051 BP.
XX
XX
AC AAT72685;
XX
XX
DT 18-SEP-1997 (first entry)
XX
XX      Sugar biosynthesis gene cluster.
XX      DE
XX      Polyketide; glycosylation; eryBIV; eryBV; eryCVI; eryBVI; eryCIV;
XX      KW eryCV; eryBVII; L-mycarose; D-desosamine; antimicrobial; antibiotic;
XX      KW antifungal; fungicide; anticancer; cytostatic; anthelmintic; ss.
XX      XX
XX      Saccharopolyspora erythraea.
XX      OS
XX      Key      Location/Qualifiers
XX      FH      CDS      80..1048
XX      FT      /*tag= a
XX      FT      /label= eryBIV
XX      FT      1048..2295

```

```

FT      /*tag= b
FT      /label= eryBV
FT      2348..3061
FT      /*tag= c
FT      /label= eryCVI
FT      3214..4677
FT      /*tag= d
FT      /label= eryHVI
FT      4674..5879
FT      /*tag= e
FT      /label= eryCIV
FT      5917..7386
FT      /*tag= f
FT      /label= eryCV
FT      7415..7996
FT      /*tag= g
FT      /label= eryHVII
FT      XX
XX      WO9723630-A2.
XX      PN
XX      03-JUL-1997.
XX      PD
XX      23-DEC-1996; 96WO-US20238.
XX      PF
XX      21-DEC-1995; 95US-0576626.
XX      PR
XX      (ABBO ) ABBOTT LAB.
XX      PA
XX      Donadio S, Katz L, Staver MJ, Summers RG;
XX      PI
XX      WPI; 1997-351066/32.
XX      DR
XX      P-PSDB; AAW19737-42.
XX      XX
XX      New genes involved in sugar biosynthesis and attachment - used to
XX      PT generate polyketide antimicrobials etc. with altered pattern of
XX      PT glycosylation
XX      PT
XX      Claim 1; Fig 4B; 85pp; English.
XX      PS
XX      2. Polynucleotides (AAT72684 and AAT72685) represent 2 distinct clusters
XX      CC of genes encoding enzymes (AAW19734-36 and AAW19737-42) involved in
XX      CC sugar biosynthesis and attachment in Saccharopolyspora erythraea.
XX      CC The eryB genes are involved in the biosynthesis of L-mycarose and
XX      CC the eryC genes involved in the biosynthesis of D-desosamine.
XX      CC Novel glycosylation-modified polyketides are produced by
XX      CC selectively altering, inactivating or augmenting these eryB and/or
XX      CC eryC genes and introducing them into polyketide-producing
XX      CC microorganisms.
XX      XX
XX      Sequence 8051 BP; 1225 A; 2898 C; 2765 G; 1163 T; 0 other;
XX      SQ

Query Match      4.0%; Score 18; DB 18; Length 8051;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 acgacgggctgctgccc 436
Db 5311 acgacgggctgctgccc 5328

RESULT 13
AAX25774
ID AAX25774 standard; cDNA; 8160 BP.
XX
XX      AAX25774;
XX      AC
XX      08-JUN-1999 (first entry)
XX      DT
XX      S.erythraea erythromycin-synthesis gene cluster eryAI-eryK.
XX      DE
XX      Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
XX      KW secondary metabolite; eryBII; eryCIII; eryCII; hybridisation; probe;
XX      KW

```

glycosylation; macrolactone; oleandomycin; ds.

Saccharopolyspora erythraea.

Key Location/Qualifiers

CDS 242..1210

FT /*tag= a

FT /label= ORF13

FT /gene= "eryBIV"

FT /product= "dTDP-keto-L-6-deoxyhexose-4-reductase"

FT 1210..2457

FT /*tag= b

FT /label= ORF14

FT /gene= "eryBV"

FT /product= "mycarosyltransferase"

FT 2510..3223

FT /*tag= c

FT /label= ORF15

FT /gene= "eryCVI"

FT /product= "dTDP-D-6-deoxyhexose-3-N-methyltransferase"

FT 3308..4840

FT /*tag= d

FT /label= ORF16

FT /gene= "eryBVI"

FT /product= "dTDP-4-keto-L-6-deoxyhexose-2,3-dehydratase"

FT 4837..6042

FT /*tag= e

FT /label= ORF17

FT /gene= "eryCIV"

FT /product= "dTDP-D-6-deoxyhexose-3,4-dehydratase"

FT 6080..7549

FT /*tag= f

FT /label= ORF18

FT /gene= "eryCV"

FT /product= "dTDP-D-4,6-dideoxyhexose-3,4-reductase"

FT 7578..8159

FT /*tag= g

FT /label= ORF19

FT /gene= "eryBVII"

FT /product= "dTDP-4-keto-D-6-deoxyhexose-3,5-epimerase"

XX W09905283-A2.

FN 04-FEB-1999.

PD 98WO-FR01593.

XX 21-JUL-1998;

XX 12-JUN-1998; 98FR-0007411.

PR 25-JUL-1997; 97FR-0009458.

XX (HMRI) HOECHST MARION ROUSSEL.

XX Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC;

PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;

XX WPI: 1999-142938/12.

DR P-PSDB: AAW99338, AAW99390, AAW99391, AAW99392, AAW99393, AAW99394,

DR AAW99395.

XX New nucleic acid sequences encoding enzymes involved in macrolide

PT biosynthesis - useful for producing hybrid secondary metabolites,

PT particularly erythromycin analogues

XX Claim 8: Fig 3: 221pp; French.

XX This sequence represents the eryAI-eryK gene cluster from the

CC Gram-positive bacterium Saccharopolyspora erythraea which encodes

CC enzymes involved in the production of the macrolide antibiotic

CC erythromycin as a secondary metabolite. The erythromycin gene cluster

CC spans approximately 53 kb and contains at least 20 open reading frames

CC (ORF). This sequence contains the eryBIV, eryBV, eryCVI, eryBVI,

CC eryCIV, eryCV and eryVII genes encoding ORFs 13, 14, 15, 16, 17, 18 and

CC 19 respectively. A DNA representing the eryG-eryAII1 region of the same

CC cluster is shown in AAX25772. The genes are used to produce hybrid

CC secondary metabolites in *S.erythraea*, i.e. erythromycin analogues which

CC may have improved properties or as hybridisation probes for isolating

CC homologous genes involved in glycosylation of macrolactones in

CC macrolide-producing strains (specifically oleandomycin-producing strains

CC of *Streptomyces antibioticus*).

XX Sequence 8160 BP; 1259 A; 2927 C; 2789 G; 1185 T; 0 other;

SQ

Query Match 4.0%; Score 18; DB 20; Length 8160;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 acgacgggctgcctgacgcg 436

Db 5474 acgacgggctgcctgacgcg 5491

|||||

RESULT 14

AAA64890/C

ID AAA64890 standard; DNA; 35026 BP.

XX AAA64890;

XX 02-FEB-2001 (first entry)

DE Bordetella pertussis pathogenicity island coding sequence.

XX

KW Bacterial infection; anti-bacterial; vaccine; whooping cough;

KW type III secretion system; virulence factor; pathogenicity island; ss.

XX Bordetella pertussis.

XX Key Location/Qualifiers

FT 711..2024

FT CDS /*tag= a

FT /product= "Protein # 1 (AAB14147)"

FT complement (2055..3590)

FT /*tag= b

FT /product= "Protein # 2 (AAB14148)"

FT 4220..4696

FT /*tag= c

FT /product= "Protein # 3 (AAB14149)"

FT complement (4998..5948)

FT /*tag= d

FT /product= "Protein # 4 (AAB14150)"

FT 6160..6747

FT CDS /*tag= e

FT /product= "Orf1 (AAB14132)"

FT complement (6783..7049)

FT /*tag= f

FT /product= "Bscf (AAB14116)"

FT complement (7049..7338)

FT /*tag= g

FT /product= "Bscf (AAB14115)"

FT complement (7374..8659)

FT /*tag= h

FT /product= "Bscf (AAB14114)"

FT complement (8656..10755)

FT /*tag= i

FT /product= "BcrD (AAB14111)"

FT complement (10752..11120)

FT /*tag= j

FT /product= "Orf2 (AAB14133)"

FT complement (11117..11527)

FT /*tag= k

FT /product= "Orf3 (AAB14134)"

FT complement (11532..11909)

FT /*tag= l

FT /product= "Orf4 (AAB14135)"

FT complement (11906..13003)

FT /*tag= m

FT CDS /product= "BopN (AAB141131)"
FT 13002..13784
FT /*tag= n
FT /product= "Orf5 (AAB141136)"
FT 13806..14081
FT /*tag= o
FT /product= "Orf6 (AAB141137)"
FT 14097..14582
FT /*tag= p
FT /product= "BcrH (AAB141112)"
FT 14630..15571
FT /*tag= q
FT /product= "Orf7 (AAB141138)"
FT 15601..16803
FT /*tag= r
FT /product= "Orf8 (AAB141139)"
FT 16827..17288
FT /*tag= s
FT /product= "Orf9 (AAB141140)"
FT 17293..17814
FT /*tag= t
FT /product= "Orf10 (AAB141141)"
FT 17892..18218
FT /*tag= u
FT /product= "BscI (AAB141117)"
FT 18215..19039
FT /*tag= v
FT /product= "BscJ (AAB141118)"
FT 19032..19694
FT /*tag= w
FT /product= "BscK (AAB141119)"
FT 19664..20302
FT /*tag= x
FT /product= "BscL (AAB141120)"
FT 20307..21641
FT /*tag= y
FT /product= "BscN (AAB141121)"
FT 21641..22150
FT /*tag= z
FT /product= "BscO (AAB141122)"
FT 22147..22695
FT /*tag= aa
FT /product= "BscP (AAB141123)"
FT 22692..23771
FT /*tag= ab
FT /product= "BscQ (AAB141124)"
FT 23768..24439
FT /*tag= ac
FT /product= "BscR (AAB141125)"
FT 24445..24711
FT /*tag= ad
FT /product= "BscS (AAB141126)"
FT 24723..25523
FT /*tag= ae
FT /product= "BscT (AAB141127)"
FT 25520..26569
FT /*tag= af
FT /product= "BscU (AAB141128)"
FT 26566..26964
FT /*tag= ag
FT /product= "BscV (AAB141129)"
FT 26955..28757
FT /*tag= ah
FT /product= "BscC (AAB141113)"
FT complement (28778..29380)
FT /*tag= ai
FT /product= "BrpL (AAB141130)"
FT complement (29412..29591)
FT /*tag= aj
FT /product= "Orf11 (AAB141142)"
FT complement (29555..30529)
FT /*tag= ak
FT /product= "Orf12 (AAB141143)"

FT CDS 30631..31776
FT /*tag= al
FT /product= "Orf13 (AAB141144)"
FT complement (31773..33005)
FT /*tag= am
FT /product= "Orf14 (AAB141145)"
FT 32370..33014
FT /*tag= an
FT /product= "Orf15 (AAB141146)"
FT complement (33002..34852)
FT /*tag= ao
FT /product= "Protein # 5 (AAB141151)"
XX WO200037493-A2.
PN 29-JUN-2000.
PD
XX 21-DEC-1999; 99WO-EPI0297.
XX 21-DEC-1998; 98GB-0028217.
XX (ULBR) UNIV LIBRE BRUXELLES.
XX Bollen A, Fauconnier A, Godfroid E;
XX WPI; 2000-452178/39.
DR P-PSDB; AAB14111, AAB14112, AAB14113, AAB14114, AAB14115, AAB14116,
DR AAB14117, AAB14118, AAB14119, AAB14120, AAB14121, AAB14122, AAB14123,
DR AAB14124, AAB14125, AAB14126, AAB14127, AAB14128, AAB14129, AAB14130,
DR AAB14131, AAB14132, AAB14133, AAB14134, AAB14135, AAB14136,
DR AAB14137, AAB14138, AAB14139, AAB14140, AAB14141, AAB14142, AAB14143,
DR AAB14144, AAB14145, AAB14146, AAB14147, AAB14148, AAB14149, AAB14150, AAB14151.
XX
PT Novel polypeptides derived from Bordetella pertussis, useful for
PT treating and diagnosing Bordetella infection -
XX
XX Example 2; Fig 5; 165pp; English.
XX
CC Bordetella pertussis possesses a type III secretion system. Type III
CC secretion systems allow bacteria to target virulence factors directly at
CC host cells. A pathogenicity island is a compact, distinct genetic unit
CC carrying virulence genes. The present sequence encodes a pathogenicity island
CC from B. pertussis. The present sequence encodes a number of proteins
CC involved in the type III secretion system of B. pertussis i.e. Bordetella
CC pathogenicity protein. The proteins encoded by the present sequence may
CC be used to treat or diagnose B. pertussis infection e.g. as a vaccine.
CC Whooping cough is a disease caused by infection by B. pertussis.
XX
SQ Sequence 3502a BP; 5572 A; 11006 C; 12284 G; 6164 T; 0 other;

Query Match 4.0%; Score 18; DH 21; Length 35026;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 ccgacgcgcacrtccgcg 187
|||||
Db 2998 CCGACGCGCGACCTCACC 2981

RESULT 15
AAH68526
ID AAH68526 standard; DNA; 349980 BP.
XX
AC AAH68526;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7061.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.

Search completed: April 3, 2002, 02:03:22
Job time: 3311 sec


```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464.052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-464-052-1

Query Match          38.9%; Score 176; DB 3; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcgggcacacccgttgaagcgtgtagcgaactgccccgcgcgcacaccccg 337
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1106 CGCGGCCACCTTTGAAGCGTGTAGCAAACTGGCGCCCCCGCGCGACCAACCG 1165

QY 338 acgacacaccccggtcatcgacacacaccccgatgcggcgccatcgacgcacccc 397
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1166 AGGACACACCCCGGTTCATCGACACACCCCGATGCGCGCGCCATCGACCGGACACC 1225

QY 398 gcagcgaaccccaacgcacacacgcggctgctgccccggctgcgcgcgtgalc 453
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1226 GCAGCCAAAGCCCAACGACACCAACGAGCGGCTCTGGCGGGGCTGCGCGCCTGATC 1281

RESULT 5
US-08-461-002-1
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485H)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
```

```
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-1

Query Match          38.9%; Score 176; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcgggcacacccgttgaagcgtgtagcgaactgccccgcgcgcacaccccg 337
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1106 CGCGGCCACCTTTGAAGCGTGTAGCAAACTGGCGCCCCCGCGCGACCAACCG 1165

QY 338 acgacacaccccggtcatcgacacacaccccgatgcggcgccatcgacgcacccc 397
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1166 AGGACACACCCCGGTTCATCGACACACCCCGATGCGCGCGCCATCGACCGGACACC 1225

QY 398 gcagcgaaccccaacgcacacacgcggctgctgccccggctgcgcgcgtgalc 453
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1226 GCAGCCAAAGCCCAACGACACCAACGAGCGGCTCTGGCGGGGCTGCGCGCCTGATC 1281

RESULT 6
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.40
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 546
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-689-411-1

Query Match          38.9%; Score 176; DB 4; Length 1535;
```



```
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 1106 CGCGGGCCACCTTTGAAGCGGTGTAGCCAAACTGGCGGCCCGCGCGGCGGACCAACCCG 1165
QY 338 acgaccacaccccggtcatcgacaccccccgcgctgctggcgcccatcgaccgacaccc 397
Db 1166 AGGACCACACCCCGTATCGACACACCCCGCGGCGGCGCATCGACCGGACACCC 1225
QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgctggcggggtcgcgcgctgac 453
Db 1226 GCAGCAAGCCCAACGCAACACGACGCGGTGCTGGCGGGGCTGCGCGGCGTGTATC 1281

RESULT 7
PCT-US94-09863-1
; Sequence 1, Application PC/TUS9409863
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
; UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael L. Goldman
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy-disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman Mr., Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1000
; TELEFAX: (716)-263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1

Query Match 38.9%; Score 176; DB 5; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 1106 CGCGGGCCACCTTTGAAGCGGTGTAGCCAAACTGGCGGCCCGCGCGGCGGACCAACCCG 1165
QY 338 acgaccacaccccggtcatcgacacacaccccgatcgccgcccacacccgacaccc 397
Db 1166 AGGACCACACCCCGTATCGACACACCCCGCGGCGGCGCATCGACCGGACACCC 1225
QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgctggcggggtcgcgcgctgac 453
Db 1226 GCAGCAAGCCCAACGCAACACGACGCGGTGCTGGCGGGGCTGCGCGGCGTGTATC 1281
```

```
RESULT 8
US-08-390-878-18
; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Hourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 38.9%; Score 176; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 9.5e-75;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 747 CGCGGGCCACCTTTGAAGCGGTGTAGCCAAACTGGCGGCCCGCGGCGGACCAACCCG 796
QY 338 acgaccacaccccggtcatcgacacacaccccgatcgccgcccacacccgacaccc 397
Db 797 AGGACCACACCCCGTATCGACACACCCCGCGGCGGCGCATCGACCGGACACCC 856
QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgctggcggggtcgcgcgctgac 453
Db 857 GCAGCAAGCCCAACGCAACACGACGCGGTGCTGGCGGGGCTGCGCGGCGTGTATC 912

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
```


COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/982,712
FILING DATE: 19921127
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-07-982-712-4

Query Match 3.8%; Score 17; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 gactggctacacccga-173
|||||
DB 7 GACTGGCTACACCCGA 23

RESULT 13
US-08-081-072-18
Sequence 18, Application US/08081072
Patent No. 5641654
GENERAL INFORMATION:
APPLICANT: No. 5641654oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS
TITLE OF INVENTION: DIAGNOSIS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, Bronstein, Roberts & Cushman
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-4280
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50inch, 1.4Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh Classic
SOFTWARE: Microsoft Word Version 4.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,072
FILING DATE: June 22, 1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 742 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
US-08-081-072-18

Query Match 3.8%; Score 17; DB 1; Length 742;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 caatacgcacggcatgctc 247
|||||
DB 358 CAATACGACGGCATGTC 374

RESULT 14
US-08-449-093A-18
Sequence 18, Application US/08449093A
Patent No. 5662906
GENERAL INFORMATION:
APPLICANT: No. 5662906oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS
TITLE OF INVENTION: DIAGNOSIS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, Bronstein, Roberts & Cushman, LLP
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-4280
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSPQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,093A
FILING DATE: May 24, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/081,072
FILING DATE: June 22, 1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/726,141
FILING DATE: July 8, 1991
CLASSIFICATION: 424
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 742 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
US-08-449-093A-18

Query Match 3.8%; Score 17; DB 1; Length 742;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 caatacgcacggcatgctc 247
|||||
DB 358 CAATACGACGGCATGTC 374

RESULT 15
US-08-081-072-15
Sequence 15, Application US/08081072
Patent No. 5641654
GENERAL INFORMATION:
APPLICANT: No. 5641654oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS

Wed Apr 3 08:33:14 2002

```

; TITLE OF INVENTION: DIAGNOSIS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-4280
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.50inch, 1.4Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh Classic
; SOFTWARE: Microsoft Word Version 4.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081,072
; FILING DATE: June 22, 1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 932 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to genomic RNA
;
US-08-081-072-15

```

```

Query Match      3.8%; Score 17; DB 1; Length 932;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 231 caatacagcggcatgtc-247
    |||||||
Db 424 CAATACGCGCATGTC 440

```

Search completed: April 3, 2002, 03:06:30
Job time: 8784 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM nucleic - nucleic search, using sw model
Run on: April 2, 2002, 22:46:21 ; Search time 1557.61 Seconds
(without alignments)
3125.197 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccgggtgcgcgcgtgatc 453

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 11351937 seqs, 5372889281 residues
Word size : 0
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlir:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	20	4.4	1100	11	BG285414 602409782
2	19	4.2	454	11	BF202807 WHE1784.H
3	19	4.2	458	11	BF924284 RC2-NT022
4	19	4.2	477	13	AQ573457 nbxb0081E
5	19	4.2	494	10	AW287746 LCL_271.H
6	19	4.2	525	10	BE499331 WHE0973.F
7	19	4.2	563	10	BE490055 WHE0364.A
8	19	4.2	575	10	BE416425 MUG008.E0
9	19	4.2	581	10	AA949237 LD28056.5
10	19	4.2	583	10	AA949237 LD28056.5
11	19	4.2	598	11	BF164631 RF04343.5
12	19	4.2	626	10	A1518445 1037958.5

13	19	4.2	650	11	BI168066
14	19	4.2	742	10	AA941556
15	19	4.2	761	13	AQ330146
16	19	4.2	802	10	AA440332
17	19	4.2	874	13	CNS01X7V
18	19	4.2	1163	13	BI10564
19	19	4.2	1273	13	AQ365308
20	18	4.0	256	11	BE850262
21	18	4.0	308	11	BF804832
22	18	4.0	355	11	BI135353
23	18	4.0	441	10	BE490643
24	18	4.0	446	10	AU091937
25	18	4.0	480	11	BF083249
26	18	4.0	483	11	BF083249
27	18	4.0	491	10	AW076894
28	18	4.0	519	11	BI142385
29	18	4.0	539	10	BE290933
30	18	4.0	554	13	AQ398116
31	18	4.0	577	10	AA654704
32	18	4.0	587	10	BE270509
33	18	4.0	638	10	AJ274323
34	18	4.0	670	11	BG810322
35	18	4.0	781	13	AQ848738
36	18	4.0	789	11	BF699432
37	18	4.0	796	11	BF268065
38	18	4.0	809	10	BE196478
39	18	4.0	856	13	CNS01XEP
40	18	4.0	857	11	BF026228
41	18	4.0	871	13	CNS04PLX
42	18	4.0	902	13	BI132519
43	18	4.0	911	11	BF784991
44	18	4.0	929	13	CNS01IWM
45	18	4.0	1009	10	AL534377

ALIGNMENTS

RESULT 1

BG285414 1100 bp mRNA EST 21-FEB-2001
LOCUS 602409782F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539236 5',
DEFINITION mRNA sequence.

ACCESSION BG285414
VERSION BG285414.1 GI:13037347

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 1100)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Arrayed by: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10467 row: d column: 21

High quality sequence stop: 739.

Location/Qualifiers

FEATURES

1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4539236"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 255 a 330 c 280 g 232 t 3 others
ORIGIN

Query Match 4.4%; Score 20; DB 11; Length 1100;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 gacaccaccccgatgcgc 377
|||||
Db 967 GACACCACCCGATGCGC 986

RESULT 2

LOCUS BF202807 454 bp mRNA EST 06-NOV-2000
DEFINITION WHE1784_H02_004Zs wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE1784_H02_004, mRNA sequence.
ACCESSION BF202807
VERSION BF202807.1 GI:111117549
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
1 (bases 1 to 454)
REFERENCE Anderson,O.D.,Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
CONTACT: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1..454
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1784_H02_004"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phuescript phagemids in the T7 Close lab (Choi, Close, Fenton) at the University of California, Riverside. plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 94 a 173 c 135 g 52 t
ORIGIN

Query Match

4.2%; Score 19; DB 11; Length 454;

Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgtacgccccagcggtc 153
|||||
Db 157 CCGCTACGCCACGGGTC 175

RESULT 3

LOCUS BF924284 458 bp mRNA EST 19-JAN-2001
DEFINITION RC2-NT0225-271100-011-h09 NT0225 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF924284
VERSION BF924284.1 GI:12320172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 458)
REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&L2-RC2-NT0225-271100-011-h09&t3=2000-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 30
High quality sequence stop: 90.
Location/Qualifiers
1..458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH25"
/dev_stage="Adult"

BASE COUNT 97 a 131 c 140 g 90 t
ORIGIN

FEATURES
Source

1..458
/note="Organ: human tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
Location/Qualifiers
97 a 131 c 140 g 90 t

BASE COUNT
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 458;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 caaccacgacgactgctg 432
|||||
Db 85 CAACACGACGGCTGCTG 103

RESULT 4
AQ573457

Fax: 5105595818

Email: oandersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

source

1. .525
Location/Qualifiers
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0973_F08_L15"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site1: EcoRI; Site2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT

111 a 200 c 153 g 61 t

ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||
Db 161 CCGCTAGCCCGCGGTC 179

RESULT 7

BE490055 563 bp mRNA EST 31-JUL-2000
LOCUS WHE0364_A04_A08ZS Wheat cold-stressed seedling cDNA library
DEFINITION Triticum aestivum cDNA clone WHE0364_A04_A08, mRNA sequence.
ACCESSION BE490055
VERSION BE490055.1 GI:9609588
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 563)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Cold-stressed seedling cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1. .563
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"

FEATURES

source

/clone="WHE0364_A04_A08"
/clone_lib="Wheat cold-stressed seedling cDNA library"
/tissue_type="Seedling"
/dev_stage="Five-day old seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site1: EcoRI; Site2: XhoI; Seeds were surface-sterilized , germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and cotopaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 110 a 206 c 163 g 84 t

ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 563;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||
Db 154 CCGCTAGCCCGCGGTC 172

RESULT 8

BE416425 575 bp mRNA EST 24-JUL-2000
LOCUS MUG008_E09R90628 ITEC MUG Wheat Spikelet Library Triticum aestivum
DEFINITION CDNA clone MUG008_E09, mRNA sequence.
ACCESSION BE416425
VERSION BE416425.1 GI:9414271
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 575)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Lanrrique,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Quilset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Watkinson,M. and Wenzel,G.
TITLE International Triticale EST Cooperative (ITEC): Production of Expressed Sequence Tags for species of the Triticeae Unpublished (2000)
JOURNAL Contact: Ogihara Y
COMMENT Kihara Institute for Biological Research, Yokohama City University
Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .575
/organism="Triticum aestivum"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone="MUG008_E09"
/clone_lib="ITEC MUG Wheat Spikelet Library"
/tissue_type="young spikelets"
/dev_stage="Feekes' scale 6-7"
/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2: XhoI; M13 Reverse sequencing primer used. 1.2 kbp average

FEATURES

source

Wed Apr 3 08:33:17 2002

```

BASE COUNT      123 a      199 c      167 g      82 t      4 others
ORIGIN
    insert size."
Query Match      4.2%; Score 19; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||
Db 173 CCGCTACGCCAGCGGGTC 191

RESULT 9
BE416426      591 bp      mRNA      24-JUL-2000
LOCUS      MUG008.E10R990628 ITEC MUG Wheat Spikelet Library Triticum aestivum
DEFINITION      cDNA clone MUG008.E10, mRNA sequence.
ACCESSION      BE416426
VERSION      BE416426.1 GI:9414272
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M. and Wenzel, G.
TITLE      International Triticeae EST Cooperative (ITEC): Production of
JOURNAL      Expressed Sequence Tags for Species of the Triticeae
COMMENT      Unpublished (2000)
Contact: Ogihara Y
Kihara Institute for Biological Research, Yokohama City University
Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES
    Source
    1..581
    /organism="Triticum aestivum"
    /cultivar="Norin 26"
    /db_xref="taxon:4565"
    /clone_lib="MUG008.E10"
    /tissue_type="young spikelets"
    /dev_stage="Feekes' scale 6-7"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; M13 Reverse sequencing primer used. 1.2 kbp average
    insert size."
BASE COUNT      127 a      199 c      167 g      82 t
ORIGIN
    insert size."
Query Match      4.2%; Score 19; DB 10; Length 581;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||
Db 173 CCGCTACGCCAGCGGGTC 191

RESULT 10
AA949237      583 bp      mRNA      EST
LOCUS

```

```

DEFINITION      LD28056.5prime LD Drosophila melanogaster embryo p0R2 Drosophila
melanogaster cDNA clone LD28056 5prime, mRNA sequence.
ACCESSION      AA949237
VERSION      AA949237.1 GI:3111146
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 583)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsunoda, G.,
Lewis, S. and Rubin, G.M.
BDGP/HIMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 280 row: E column: 8
High quality sequence stop: 400.
Location/Qualifiers
    1..583
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="LD28056"
    /clone_lib="LD Drosophila melanogaster embryo p0R2"
    /sex="male and female"
    /dev_stage="0 to 24 hours mixed stage embryonic"
    /lab_host="XLI Blue"
    /note="organ: embryo; Vector: p0R2; Site_1: EcoRI; Site_2:
    XhoI; Sized fractionated cDNAs were directly ligated into
    p0R2."
BASE COUNT      151 a      181 c      151 g      100 t
ORIGIN
    Query Match      4.2%; Score 19; DB 10; Length 583;
    Best Local Similarity 100.0%; Pred. No. 69;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccagcagcggtcgtggcc 435
|||||
Db 543 CCACGACGGGCTGTGGCC 561

RFSUIT 11
B1164631
LOCUS      B1164631      598 bp      mRNA      EST      09-JUL-2001
DEFINITION      RE0434.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE04343 5 similar to CG4532;
Rhan0004532 located on: X 6D1-6D2; 04/11/2001, mRNA sequence.
ACCESSION      B1164631
VERSION      B1164631.1 GI:14630437
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 598)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champ, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celnik, S. and Rubin
G.M.
BDGP/HIMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

```

```

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AF003438: arm:X [6347108,6646644]
estimated-cyto:6B1-6D3: 04/11/2001
Plate: RE.43 row: D column: 7
High quality sequence stop: 490.
Location/Qualifiers
1. .598
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFLC-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pFLC1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 156 a 185 c 157 g 99 t 1 others
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 598;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
DB 567 CCACGACGGGCTGCTGGCC 585

RESULT 12
LOCUS AI518445 626 bp mRNA EST 19-APR-2001
DEFINITION LD37958.5prime LD Drosophila melanogaster embryo pot2 Drosophila
melanogaster cDNA clone LD37958 5prime, mRNA sequence.
ACCESSION AI518445.1 GI:4424299
VERSION EST.
KEYWORDS fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
M.scomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 626)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/RHMI Drosophila EST Project
Unpublished (2001)
CONTACT: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 379 row: E column: 10
High quality sequence stop: 413.
Location/Qualifiers
1. .626
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD37958"
/clone_lib="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/note="Organ: embryo; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2."

```

```

BASE COUNT 165 a 192 c 162 g 107 t
ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 626;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
DB 563 CCACGACGGGCTGCTGGCC 581

RESULT 13
LOCUS BI168066 650 bp mRNA EST 09-JUL-2001
DEFINITION RE08607.5prime RE Drosophila melanogaster normalized Embryo pFLC-1
Drosophila melanogaster cDNA clone RE08607 5 similar to CG4532:
FBan0004532 located on: X 6D1-6D2; 04/11/2001, mRNA sequence.
ACCESSION BI168066
VERSION EST.
KEYWORDS fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 650)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
,G.M.
BDGP/HIMI RE Drosophila EST Project
Unpublished (2001)
CONTACT: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003438: arm:X [6347108,6646644]
estimated-cyto:6B1-6D3: 04/11/2001
Plate: RE.86 row: A column: 7
High quality sequence stop: 515.
Location/Qualifiers
1. .650
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE08607"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFLC-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pFLC1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 155 a 207 c 172 g 115 t 1 others
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 650;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
DB 589 CCACGACGGGCTGCTGGCC 607

```

RESULT 14
AA941556 742 bp mRNA EST 23-APR-2001
LOCUS
DEFINITION
LD25748.5prine LD Drosophila melanogaster embryo pOT2 Drosophila
melanogaster cDNA clone LD25748.5 similar to CG4532: FBan0004532
located on: X 6D1-6D2; 04/10/2001, mRNA sequence.

ACCESSION
AA941556
VERSION
AA941556.2 GI:13767860
KEYWORDS
EST.
SOURCE
Drosophila melanogaster
fruit fly.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscumorphia; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 742)
REFERENCE
AUTHORS
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HIMI Drosophila EST Project
Unpublished (2001)
TITLE
JOURNAL
COMMENT
On May 1, 1998 this sequence version replaced gi:3101469.
Other ESTs: LD25748.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003438: arm: X [6347108..6646644]
estimated-cyto:681-6D3: 04/10/2001
plate: LD.257 row: D column: 12
High quality sequence stop: 605
POLYA-No.

FEATURES
source
1..742
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD25748"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="X11 Blue"
/notes="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."

BASE COUNT 196 a 219 c 195 g 132 t

ORIGIN
Query Match 4.2%; Score 19; DB 10; Length 742;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgagctactgccc 435
|||||
Db 571 CCACACGGGCTGCTGCC 589

RESULT 15
A0330146 761 bp DNA GSS 08-JAN-1999
LOCUS
DEFINITION
nxb0046H15f CUGI Rice BAC Library Oryza sativa genomic clone
nxb0046H15f, DNA sequence.
A0330146
VERSION
A0330146.1 GI:4121996
KEYWORDS
GSS.
SOURCE
Oryza sativa.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 761)
REFERENCE
AUTHORS
Wing,R.A. and Dean,R.A.
TITLE
A RAC End Sequencing Framework to Sequence the Rice Genom

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:44:13 ; Search time 2977.34 Seconds
(without alignments)
2510.033 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccggcgctgcgcgcgtgatc 453

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_pl:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sv:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_on:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vt:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	415.8	91.8	3453	1	MTU43540	U43540 Mycobacteri
2	410	90.5	1604	1	MBU35021	U35021 Mycobacteri
3	410	90.5	9281	1	MBDR351	U35017 Mycobacteri
C 4	410	90.5	9764	1	AE006921	AE006921 Mycobacte
C 5	410	90.5	10019	1	AF041819	AF041819 Mycobacte
C 6	410	90.5	17783	1	AE007028	AE007028 Mycobacte
C 7	408.4	90.2	1535	6	AR096713	AR096713 Sequence
C 8	408.4	90.2	1535	6	ARI47694	ARI47694 Sequence
C 9	408.4	90.2	32437	1	MTCY336	295586 Mycobacteri
C 10	408.4	90.2	38380	1	MTCY251	274410 Mycobacteri
C 11	408.4	90.2	14401	1	MTY13E12	295390 Mycobacteri
C 12	403.6	89.1	16384	1	AF007160	AE007160 Mycobacte
C 13	366.8	81.0	12412	6	186264	186264 Sequence 18
C 14	351	77.5	650	6	AR096715	AR096715 Sequence
C 15	351	77.5	650	6	AR147696	AR147696 Sequence
C 16	214.6	47.4	14179	1	AE007053	AE007053 Mycobacte
C 17	214.6	47.4	33100	1	MTCY9H9	284498 Mycobacteri
C 18	205	45.3	34331	1	MTY165	295584 Mycobacteri
C 19	203.4	44.9	14953	1	AE006996	AE006996 Mycobacte
C 20	189.2	41.8	40056	1	MSGY348	AD000020 Mycobacte
C 21	183.8	40.6	15019	1	AE007036	AE007036 Mycobacte
C 22	183.8	40.6	37432	1	MTY1125	298268 Mycobacteri
C 23	137.4	30.3	15504	1	AE006995	AE006995 Mycobacte
C 24	135.8	30.0	22550	1	MTCY22G8	295585 Mycobacteri
C 25	63.8	14.1	43254	1	MLCBI779	298271 Mycobacteri
C 26	63.8	14.1	312050	1	MLEPRTN3	AL583919 Mycobacte
C 27	60.6	13.4	348450	1	MLEPRTN4	AL583920 Mycobacte
C 28	59	13.0	819	11	CNS06EWP	AL395663 T7 end of
C 29	57.2	12.6	64957	1	AB032367	AB032367 Streptomy
C 30	56	12.4	177883	2	AC046159	AC046159 Homo sapi
C 31	55.6	12.3	11096	1	AF275943	AF275943 Streptomy
C 32	55.6	12.3	12381	6	AX006889	AX006889 Sequence
C 33	55.6	12.3	64957	1	AB032367	AB032367 Streptomy
C 34	55.2	12.2	4231	8	SVHKGPG	X56010 Sorghum vul
C 35	54.2	12.0	197669	2	AC084064	AC084064 Homo sapi
C 36	54.2	12.0	265537	2	AC087228	AC087228 Mus muscu
C 37	54	11.9	56870	2	CXY39B6_3	Continuation (4 of
C 38	53.8	11.9	69173	2	AC025052	AC025052 Homo sapi
C 39	53.6	11.8	82024	2	AC023210	AC023210 Homo sapi
C 40	53.4	11.8	840	8	CNS0191G	AL111680 BoLyris
C 41	53.4	11.8	62649	2	AC022552	AC022552 Homo sapi
C 42	53.4	11.8	110737	2	AC011105	AC011105 Homo sapi
C 43	53.4	11.8	205914	2	AC027682	AC027682 Homo sapi
C 44	53.2	11.7	78220	2	AC023212	AC023212 Homo sapi
C 45	53.2	11.7	89994	2	AC021263	AC021263 Homo sapi

ALIGNMENTS

RESULT 1	MTU43540	3453 bp	DNA	BCT	14-AUG-1997
LOCUS	Mycobacterium tuberculosis rfbA, rhamnose biosynthesis protein (rfa), and rmlC genes, complete cds.				
DEFINITION	U43540				
ACCESSION	U43540.1	GI:2326948			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
	Mycobacterium tuberculosis.				
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;				
	Mycobacterium; Mycobacterium tuberculosis complex.				
REFERENCE	1 (sites)				
AUTHORS	Lee, J.				
TITLE	Rhamnose biosynthetic genes related to a novel repeated sequence of Mycobacterium tuberculosis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3453)				
AUTHORS	Lee, T.				

Matches	447;	Conservative	0;	Mismatches	5;	Indels	4;	Gaps	3;
QY	1	gacgggagggccacatacaaaagtattcgcccttttttgcgcacacctgcccgcggc							60
Db	451	GATCGCGAGGCGGCACATCAAAAGTATTCGGCCCTTTTTCGCCCTTCCTTCGCGCGG							510
QY	61	tggatgttcacacccgagggcgccgaagcgagcctcgccgcaaaagc--ctcaatata							118
Db	511	TGGATGTGTCCACCGCGCCAGCGCGCGCAAGCGGACCTTCGCGCAAGCCGCTTAATATC							570
QY	119	gtcccaacagctgcccgtacgcccgaagcggttgatgactggctacaccccgccgcg							178
Db	571	GTCCCGAGAGAGCTGGCGGCTAGCGCCAGCGCGCGTACGCTACGCTACGCTACGCTACG							630
QY	179	acctaccgacacccgagcgcccgcccaaacgc-gcatcaccttgagcaacgaatatac							237
Db	631	ACCTACCCGACACCGAAGCGCGCCGCAAAAGCGGATCACCTGAGCAACGACGAATACG							690
QY	238	acggatgtcaaggctaaagtctacctgaccccccaagtcgcccgcacaccccttgaagc							297
Db	691	ACGGCATGTACAGGGTAAGTGGCTACCTGACCCGCCAAG-CGCGGGCCACCTTTGAAGCC							749
QY	298	gtgttagcaaaactggccgcccgcgagcaaaccccgagacacaccccggtatc							357
Db	750	GTGCTAGCCAAACTGGCGCCCGCGCGGCGGACCAACCCGAGCACACCCGCGGTCAATC							809
QY	358	gacacacccccgatgcccgcgcacatcgacccgagacaccccgagcccaacgcaac							417
Db	810	GACACACCCCGATCGCGCGCCATCGACCGCGACACCGCGAGCAAGCCCAACGCAAC							869
QY	418	cacgacggctgctgcccggctgcccgcctgctgac							453
Db	870	CACGACGGCTGCTGGCGGGCTGCGCGCGCTGATC							905
RESULT	4	AE006921/C							
LOCUS		9764 bp	DNA	BCT					27-APR-2001
DEFINITION		Mycobacterium tuberculosis (DC1551, section 7 of 280 of the complete genome.							
ACCESSION		AE006921	AF000516						
VERSION		U35017.1	GI:14879142						
KEYWORDS		Mycobacterium tuberculosis CDC1551.							
SOURCE		Mycobacterium tuberculosis CDC1551.							
ORGANISM		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.							
REFERENCE		1 (bases 1 to 9764)							
AUTHORS		Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Berry, K., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., and Bishai, W.							
TITLE		Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains							
JOURNAL		Unpublished							
REFERENCE		2 (bases 1 to 9764)							
AUTHORS		Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Berry, K., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., and Bishai, W.							
TITLE		Direct Submission							
JOURNAL		Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA							
FEATURES		Location/Qualifiers							
source		1..9764							
		/organism="Mycobacterium tuberculosis CDC1551"							
		/strain="CDC1551"							
		/db_xref="taxon:83331"							
		/note="clinical strain"							
		122..1414							
		/gene="MT0095"							
		/note="This region contains an authentic point mutation,							
BASE COUNT		1706 a	3191 c	2929 g	1453 t	2	others		
ORIGIN									
Query Match		90.5%; Score 410; DB 1; Length 9281;							
Best Local Similarity		98.0%; Pred. No. 2.2e-56;							
Matches	447;	Conservative	0;	Mismatches	5;	Indels	4;	Gaps	3;

RESULT 3
 LOCUS MBD3S1 9281 bp DNA BCT 27-APR-1996
 DEFINITION Mycobacterium bovis deletion region 3, 5' end.
 ACCESSION U35017
 VERSION U35017.1 GI:1049238
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium bovis.
 Mycobacterium bovis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
 REFERENCE 1 (bases 1 to 9281)
 AUTHORS Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., and Stover, C.K.
 TITLE Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis
 JOURNAL J. Bacteriol. 178 (5), 1274-1282 (1996)
 MEDLINE 96200095
 REFERENCE 2 (bases 1 to 9281)
 AUTHORS Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., and Stover, C.K.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology, Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
 FEATURES
 source
 Location/Qualifiers
 1..9281
 /organism="Mycobacterium bovis"
 /db_xref="taxon:1765"
 misc_difference replace(1400..9281,"")
 /note="absent in related avirulent strain M. bovis BCG"
 BASE COUNT 1706 a 3191 c 2929 g 1453 t 2 others
 ORIGIN

Query Match 90.5%; Score 410; DB 1; Length 9281;
 Best Local Similarity 98.0%; Pred. No. 2.2e-56;
 Matches 447; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

causing a premature stop, and is not the result of a sequencing artifact; similar to SP:P16431 GB:X17506 PID:41684 PID:882614 GB:U00096; identified by sequence similarity; putative"

1449..2123
/gene="MT0096"
/gene="MT0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879143"
/translation="MSYIKHAPSRVRLRQTRSTVVKRSGSLSRVRRTGDLGLAVWG GREYRAVKPTGIPQKGDMMTVVVDAGPRGRSVSEVAAPAEALFALVADPRHR ELDSGTVRGNIKVPKLVGSKFTKMKLFGLPYRITSRVTKALKEINELVECSHPLGH RWRWFESLPTLRVTETEDFHAAGAINKGLKFEYEMTGFAKSNAAAGIATLAKLSDQ YARGA"

2280..2873
/gene="MT0098"
/gene="MT0098"
/note="similar to GP:145427; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAK44320.1"
/db_xref="GI:13879144"
/translation="MDQPNANIHVDALLDAMVPLGTQCVDVCCGGILLAARLARRI PVTAVDIDAPVLRRAQTRFANAPIRWLHADIMTAELPNAGFDVAVSNAALHLEDR TALSLRGLGTPGGTLAVVTFVTPSLRNLGHLTSSWACGMANRVKGRKWEHSAPIKWP PPOTLHELSHVRALLPGACIRRLLYGRVLVTRAPV"

3002..3772
/gene="MT0099"
/gene="MT0099"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNONRNRNWEELITCGLVHYAPDDAALAAKRASTGLGE WKLCRCGDFALGGPOGRGAPEDAPLIMRGKALROAIIIRALGVERLVRLVLAALAAW AVWEFRGARGAQTALDRDLPVLRAGFKVDQMTVIHALEKALAKPSTLALITCMLA AYAVLOAEGVGWLKRWGEYFAVVATSIPLFLEVDHIAKGIITTRVVTTSINVAAY VYLLSKRLFGVRGRKAYDVERRGQLLDLEKAMLT"

3791..4162
/gene="MT0099.1"
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MRGLRVYSKVTSAFFALTTPARRAKARVNLDOVAOCRRTDEG PTLQCHQPGSALPPTAAWSQRVPATHCRPPCAPCAASALTCALEAWSWVE VRPAPRAATFSLOSRAAESR"

4206..4973
/gene="MT0100"
/gene="MT0100"
/note="similar to GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase, putative"

/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTGVICAIPQELAYLGVLDAKROQVAQILFDSQDLAHR VLLAAGMGKVNLTGLTATLLADRGCTITFTSVAGGLDELICIGDIVIADRVQIDF GLITDRLRPOGCHIPFIEPTRELGVPPVDPAVIDGRKRLDGTFLAPLTAAGGGR OPRIYGTLLTGDOYLHCETRNRLHIELGGMAVEGGAQAQICASPIIPWILVHAL SDLAGADSGVDNFNRFGVGAASSARVILRLLPVLTTAC"

5105..7390
/gene="MT0101"
/gene="MT0101"
/note="similar to SP:p46840 PID:559912; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTTAVTGEHSHASVRIQLRISGMSCSACAHREVESTINKLPGVRA AVNGTRVATITUTSEAVIAAALCOAVRRAGYQADLCITDDGSRASDDPADHAROLLIKL AIAVLFPVADLTVMEGVVPATRTGKQWVLSALCLPVTWMAAPHRHMRNRHH VGRFEAKAKSOAGSALRAIALAKAKAVANLLPDGSEMWIPADFLKEQRFVVRPQI AASMETLISVGITATITNSLYTVEGNISPTERSGIMQALLGSDAIYEVAAGVTVFVL VAAUGLAVDGSAAVIMSAMTCEAKPTKVRPGQVIGGTIVLDKCLIVEAAAGADTQF ACMVKLVFQAQAKADIAURLAURISSEVFPVAVIATACMLIAGCQPDRAVSAAL AVIYIACICALCGLATITAMVASGRGAQLGIFLKGYKSLERATRAVDVTFDKYGTITTT CBLQYSVATAAGWEADQVLAIATATEAASRHSVALIAAATTRDAVTFDRAIPGRG VSGTVSGRAVRVKGPSWISGSSSCHPNMRAARHAEESLGETAFVFEVDGCEGCTAVAD AVKTSARUAAALADKGLRTMLITGDNPEAAAVATRVGIDEVIADILPEKGVDFEQ LDRHGVAMVGGVINKIPALARADLMAIGCTDVAIGAADIILVRUHLDPVPIALD LARATMTVTKGNVWAFQYNIAAIPVAAGLLNPLVAGAAAFSFFSVVNSILKKP LARATMTVTKGNVWAFQYNIAAIPVAAGLLNPLVAGAAAFSFFSVVNSILKKP GRYPGCGTGGVGGQMTAPSSA"

complement(7337..8185)
/gene="MT0102"
complement(7337..8185)
/gene="MT0102"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44325.1"
/db_xref="GI:13879149"
/translation="MLAOTATAGSFNIIHASTVLOGRGVPAAMWSEPCATIRHHCATIL DDMCEVAFALSAKILIGKRAIPVPSAKVDEHLGECSSACRAMFTQVSAQGLIKRLAES RPVVPVGRGLGIRAPKPKHSPMTWRRWALLCVGIAOIALGTQVGGFDGLVGHQHTP GACTHLLINESTSIALGVINVGAAALWPSAAGLAGVLTAFVAILTGYVIVDALSCAV STTRITLHPVWITCAVLAIMWRSASPRPRPDANAEFDIVLPDNASHGRKRLWLP TDKSAA"

complement(8232..9093)
/gene="MT0103"
complement(8232..9093)
/gene="MT0103"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44326.1"
/db_xref="GI:13879150"
/translation="MDWLHPGDLTDFERARKRGITLSNOQYDMSRLSGYLIPQARA TFEAVLAKIAPGATNPDDHTPVTDTPDAAAIIDROTTSQAQRNHDGILAGLALIAS KLGQHNGLPVSIVVTTTLLDTDTGACKGFTGGTLLPWADVIKMTSHAHYSPASGR YPQAIIPDHGTPALALYHTRKLRASPAORIMLFANDRGCTKPCCDAPAYHSAHHVTAWTS TGRTDITELTLCGPDNRLAEKGTTHKTHKTHIGHTWLPPLPHLDHGQPRNTFIHPERF LHNQDDDDKPD"

complement(8956..9623)
/gene="MT0104"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; identified by Glimmer2; putative; conserved hypothetical protein, authentic frameshift"

BASE COUNT 1553 a 3103 c 3360 g 1748 t
ORIGIN

	Query Match	90.5%; Score 410; DB 1; Length 9764;
	Best Local Similarity	98.0%; Pred. No. 2.1e-56;
	Matches 447; Conservative	0; Mismatches 5; Indels 4; Gaps 3;
Qy	1 gatcgaggagccacatcaaatgattgcggccttttgcgccactgcgccgcgcg 60 	
Db	9250 GATCGGAGGGCACAATAAGTATTGCGCCCTTTTCGCCACACTGCCCCCGGG 9191 	
Qy	61 tggatgtgtcaccccgcgcgaagcgcacctgcggcgaagc--ctcaatatc 118 	
Db	9190 TGGATGTGCCACCGCAGCGCGCGAAGCGCACTGGCGGCAACCGCTCAATATC 9131 	
Qy	119 gtcccgagagtgtgcgcgtacgccccagcgggtcatggaactggtacaccgccgcgcgc 178 	
Db	9130 GTCCCGACGAGCTGGCCGCGTACGCCACGCGGGTCATGGACTGGCTACACCCGACGCGC 9071 	
Qy	179 acctcacgacacgaacgcgcgcgcgaacgc-gcatcacctdagcaaccagcaatacgc 237 	
Db	9070 ACCTCACGACACGAACGCGCGCAACAGCGGCATCACCTGACCAACCAGCAATACG 9011 	
Qy	238 acggcatgtcacggctaagtgttcactgcaccccccaagtcgcgggcacacctttgaagcc 297 	
Db	9010 ACGCATGTCA CGGCTAA GTTGCTACCTGACCCCCCAAG -CGCGGGCCACCTTTGAAGCC 8952 	
Qy	298 gtgttaacaaactgc 357 	
Db	8951 GTGTACCAAACATGCCGCCGCCCGCGCGGACCAACCCCGACGACCAACCCCGTATC 8892 	
Qy	358 gacacaccccgatgcggcgccatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 417 	
Db	8891 GACACACCCCGATGCGCGCGCCATCGACGCGACACCCGACGCCAACGCCCAAC 8832 	
Qy	418 cacgacgggtgctgcgcgggtgc 453 	
Db	8831 CACGACGGGTGCTGCGCGGGTGC GCGCGTGCATC 8796 	
RESULT	5	
AF041819/c		
LOCUS	AF041819 10019 bp DNA BCT 18-JAN-1998	
DEFINITION	Mycobacterium bovis BCG putative	
	adenosylmethionine-8-amino-7-oxononanolate aminotransferase (bioA), putative 8-amino-7-oxononanolate synthase (bioF), putative dethiobiotin synthetase (bioD), and biotin synthetase (bioB) genes, complete cds.	
ACCESSION	AF041819	
VERSION	AF041819.1 GI:2791842	
KEYWORDS	.	
SOURCE	Mycobacterium bovis BCG.	
ORGANISM	Mycobacterium bovis BCG. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu, S. and Jacobs, W. R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019) Yu, S.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (07-JAN-1998) Microbiology & Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10010, USA	
JOURNAL	Location/Qualifiers	
FEATURES	1..10019	
source	/organism="Mycobacterium bovis BCG" /strain="Pasteur" /sub_species="BCG" /db_xref="taxon:33892" complement(259..1038) /note="ORF1; similar to MTCY336.37"	
CDS		

CDS 7468..9051
/gene="MT1631"
/note="similar to GP:608530; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AAK45899.1"
/db_xref="GI:13881260"
/translation="MAGPAWRDAADVIVIGTVAGVLAALAAADRGRSVVLSKAQT
HVTATHYAOQGGIAVLPDNDSDVDAHVDVTLAAGAGLCPDPAVYSIVADYGRVTDLV
GAGRLDSEVPGRWALTREGHSRRRIHVAGGDATGAEVORLODAAGMLDIRTHVA
LRVLHDTAVTGLLVVRPDCGIIISAPSVTLTGGLHLYSATNPAGSTGGLALGL
WAGVAYDLDFIOFHPDMLPAGRGRPLITEIRGEKAILVDRGNCNITAGVHPMG
DLAPDVAAADIALKATGDPVYLDKAGIEGFASRFPVTASCRAGGIDVPRQIP
VPGAHYSCGIVDYGTELLGLYAAAEVARTLGHANRLASNLLEGLVVGGRAG
KAAAHAAAGRSRATYSATWPEPISYALDRGLQRAMSRDASMYRAAAGLRLCDS
LSGAQVRDVACRRDFEDVALTLVAQSVTAALARTESRGCHRAEYPTVPEQARSIV
VRGADDANAVCVQALVAVC"
9021..9908
/gene="MT1632"
9021..9908
/gene="MT1632"
/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262
PID:456041; identified by sequence similarity: putative"
/codon_start=1
/transl_table=11

Query Match 90.5%; Score 410; DB 1; Length 17783;
Best Local Similarity 98.0%; Pred. No. 1.7e-56; Indels 4; Gaps 3;
Matches 447; Conservative 0; Mismatches 5;

Qy 1 gatcgagagcgcaactcaagtgattcgccctttttcgccacccctgcccgcgcg 60
Db 1125 GATCGGAGGCGCACATCAAGTATCGCGCCCTTTTTCGCCACCTGCCGCGCG 1066

Qy 61 tggatgtgtccaccgagcgccggaagcgacgacgagcgcaaggc--ctcaatalc 118
Db 1065 TGGATGTGTCCACCGCGCGGCGGCGGAGCGACCTGGCGCGCAAGCGCGTCAATATC 1006

Qy 119 gtcccgagagctggccgctacgcccagcgaggtcatgactggtctacaccccgagc 178
Db 1005 GTCCGACGAGCTGGCCGCTACGCCAGCGGTCATGACTGGCTACACCCCGAGCGCG 946

Qy 179 acctaccgacacgagcgcccgcaaacgc- gcatcaccctgagcaaccagcaatc 237
Db 945 ACCTACCGACACCGAACCGCGCGCGCAACCGCGCATCACCTGGACACACCGCAATATC 886

Qy 238 acggcatgtcagcgctaaagtgtgctacgtccgagcgaggtcatgactggtctacaccccgagc 297
Db 885 ACGGCATGTACGGCTAAGTGTACCTACCTGCCGCCCAAGCGCGCATCACCTTTGAAGCC 827

Qy 298 gtgctagcaaaactggccgccccggcgacccaccccgagcaccaccccggtc 357
Db 826 GTGCTAGCAAACTGGCGCGCGCGCGCGACCAACCCCGAGCACACCCCGGTCTATC 767

Qy 358 gacaccaccccgatcgccgcccctacgacgacaccccgacaccccgagcaccgcaac 417
Db 766 GACACACCCCGATCGCGCGCGCATCGACCGCGACACCCCGAGCCCAAGCCCAAC 707

Qy 418 cagcagggctgtgctggcggtgctgagcggtgac 453
Db 706 CACGACGGGCTGTGGCGGGGTGGCGCGGTGATC 671

RESULT 7
AR096713
LOCUS AR096713 1535 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6008201.
ACCESSION AR096713
VERSION AR096713.1 GI:10025749
KEYWORDS
SOURCE Unknown.

CDS 7468..9051
/gene="MT1631"
/note="similar to GP:608530; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AAK45899.1"
/db_xref="GI:13881260"
/translation="MAGPAWRDAADVIVIGTVAGVLAALAAADRGRSVVLSKAQT
HVTATHYAOQGGIAVLPDNDSDVDAHVDVTLAAGAGLCPDPAVYSIVADYGRVTDLV
GAGRLDSEVPGRWALTREGHSRRRIHVAGGDATGAEVORLODAAGMLDIRTHVA
LRVLHDTAVTGLLVVRPDCGIIISAPSVTLTGGLHLYSATNPAGSTGGLALGL
WAGVAYDLDFIOFHPDMLPAGRGRPLITEIRGEKAILVDRGNCNITAGVHPMG
DLAPDVAAADIALKATGDPVYLDKAGIEGFASRFPVTASCRAGGIDVPRQIP
VPGAHYSCGIVDYGTELLGLYAAAEVARTLGHANRLASNLLEGLVVGGRAG
KAAAHAAAGRSRATYSATWPEPISYALDRGLQRAMSRDASMYRAAAGLRLCDS
LSGAQVRDVACRRDFEDVALTLVAQSVTAALARTESRGCHRAEYPTVPEQARSIV
VRGADDANAVCVQALVAVC"
9021..9908
/gene="MT1632"
9021..9908
/gene="MT1632"
/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262
PID:456041; identified by sequence similarity: putative"
/codon_start=1
/transl_table=11

Query Match 90.5%; Score 410; DB 1; Length 17783;
Best Local Similarity 98.0%; Pred. No. 1.7e-56; Indels 4; Gaps 3;
Matches 447; Conservative 0; Mismatches 5;

Qy 1 gatcgagagcgcaactcaagtgattcgccctttttcgccacccctgcccgcgcg 60
Db 1125 GATCGGAGGCGCACATCAAGTATCGCGCCCTTTTTCGCCACCTGCCGCGCG 1066

Qy 61 tggatgtgtccaccgagcgccggaagcgacgacgagcgcaaggc--ctcaatalc 118
Db 1065 TGGATGTGTCCACCGCGCGGCGGCGGAGCGACCTGGCGCGCAAGCGCGTCAATATC 1006

Qy 119 gtcccgagagctggccgctacgcccagcgaggtcatgactggtctacaccccgagc 178
Db 1005 GTCCGACGAGCTGGCCGCTACGCCAGCGGTCATGACTGGCTACACCCCGAGCGCG 946

Qy 179 acctaccgacacgagcgcccgcaaacgc- gcatcaccctgagcaaccagcaatc 237
Db 945 ACCTACCGACACCGAACCGCGCGCGCAACCGCGCATCACCTGGACACACCGCAATATC 886

Qy 238 acggcatgtcagcgctaaagtgtgctacgtccgagcgaggtcatgactggtctacaccccgagc 297
Db 885 ACGGCATGTACGGCTAAGTGTACCTACCTGCCGCCCAAGCGCGCATCACCTTTGAAGCC 827

Qy 298 gtgctagcaaaactggccgccccggcgacccaccccgagcaccaccccggtc 357
Db 826 GTGCTAGCAAACTGGCGCGCGCGCGCGACCAACCCCGAGCACACCCCGGTCTATC 767

Qy 358 gacaccaccccgatcgccgcccctacgacgacaccccgacaccccgagcaccgcaac 417
Db 766 GACACACCCCGATCGCGCGCGCATCGACCGCGACACCCCGAGCCCAAGCCCAAC 707

Qy 418 cagcagggctgtgctggcggtgctgagcggtgac 453
Db 706 CACGACGGGCTGTGGCGGGGTGGCGCGGTGATC 671

RESULT 7
AR096713
LOCUS AR096713 1535 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6008201.
ACCESSION AR096713
VERSION AR096713.1 GI:10025749
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1535)
TITLE Riley, L.W.
JOURNAL DNA molecule encoding for cellular uptake of mycobacterium
FEATURES tuberculosis and uses thereof
Source Location/Qualifiers
1..1535 /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match 90.2%; Score 408.4; DB 6; Length 1535;
Best Local Similarity 97.8%; Pred. No. 8.3e-56;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

Qy 1 gatcgagagcgacatcaaaagtgtattcgccctttttcgccacccctgcccgcgcg 60
Db 827 GATCGGAGGCGCACGTCAAAGTCAATCGTCCCTTTTTCGCCACCTGCCGCGCG 886

Qy 61 tggatgtgtccaccgagcgccggaagcgacgtggtctacaccccgagc 118
Db 887 TGGATGTGTCCACCGCGCGCGCGCGGAGCGACCTGGCGCGCAAGCGCGTCAATATC 946

Qy 119 gtcccgagagctggccgctacgcccagcgaggtcatgactggtctacaccccgagc 178
Db 947 GTCCGACGAGCTGGCCGCTACGCCAGCGGTCATGACTGGCTACACCCCGAGCGCG 1006

Qy 179 acctaccgacacgagcgcccgcaaacgc- gcatcaccctgagcaaccagcaatc 237
Db 1007 ACCTCACCAGACACCGAGCGCGCGCGCAACCGCGCATCACCTGGACACACCGCAATACG 1066

Qy 238 acggcatgtcagcgctaaagtgtgctacgtccgagcgaggtcatgactggtctacaccccgagc 297
Db 1067 ACGGCATGTCAAGCTAAGTGTCTACCTGACCTGCCGCCCAAGCGCGCATCACCTTTGAAGCC 1125

Qy 298 gtgctagcaaaactggccgccccggcgacccaccccgagcaccaccccggtc 357
Db 1126 GTGCTAGCAAACTGGCGCGCGCGCGCGACCAACCCCGAGCACACCCCGGTCTATC 1185

Qy 358 gacaccaccccgatcgccgcccctacgacgacaccccgacaccccgagcaccgcaac 417
Db 1186 GACACACCCCGATCGCGCGCGCATCGACCGCGACACCCCGAGCCCAAGCCCAAC 1245

Qy 418 cagcagggctgtgctggcggtgctgagcggtgac 453
Db 1246 CACGACGGGCTGTGGCGGGGTGGCGCGGTGATC 1253

RESULT 8
AR147694
LOCUS AR147694 1535 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6224881.
ACCESSION AR147694
VERSION AR147694.1 GI:15111784
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1535)
TITLE Riley, L.W. and Chong, P.
JOURNAL DNA molecule fragments encoding for cellular uptake of
FEATURES Mycobacterium tuberculosis and uses thereof
Source Location/Qualifiers
1..1535 /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.

Gene prediction was based on a Hidden Markov Model of TH genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

```
Location/Qualifiers
1..32437
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
<1..273
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y48"
33..32091
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y336"
 /gene="Rv1565c"
 /note="Rv1565c"
 /gene="Rv1565c" (MTCV336_38), len: 729. Unknown membrane protein, some similarity to O05402 HYPOTHETICAL 72.2 KD PKPHEIN from B. subtilis (634 aa) opt: 384 E(1): 4.8e-17: 29.1% identity in 378 aa overlap and to hypothetical protein in H.influenzae. N-terminal half hydrophobic. FASTA results, Y392_UAEIN P43993 hi0392 (245 aa) opt: 255 E(1): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse score is 0.930"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1565c"
 /protein_id="CAB09072.1"
 /db_xref="GI:2117272"
 /db_xref="SPTROMBI:O06625"
 /translation="MTLSLPPRALTPPALPPVTMGTRTTGFRHDLDGLRGVALIVAVPIHWFGVRSGGVWPLALISGFPGCKILRAALNPDISPAFVIRILRRLLLIHWSWGQVFYLFLIHLVAGCATJLVGAPVPIWRPMWLKTATVAIALAILSCGGVNIHQVEFCIPMAVLVPGVATMIMLACANROGHGPDRDLPLNRLLATAPLVAIGASYSYLWHMPILFLFWLSYTGHRHANFVEGAALVLTSLGLLAYLTRLVEDPLKYRAPAKSPAAVPPLPWRLLRRTPIVLGVSVALLGVALTSTFTWREHVI VORAAGEKI.SGLSDRYFGARALIDHVRYPKLMRPRTLVEVRHDLPTSTKDGCISDFVNAPILNCTGYDAPRTITALAGGSAAEHLWTALDLGRNHFKFYTVLKMGCPILSTEPEVPLIMGNAPYLCIQWYOAMAKLVADHDIPDYFTTSTRHPWI KPGDVMPATYVGIIWTFADNNI PVLANPTPPLWKDGQPTIPADCLAKNGNQSCGTARSKVLDVRNPITLDFVARFPILKLPLDMIAICRDYTCCRVAGSNVLVYRSHSLTLPTYMTWTSELGRQIAANTDW"
```

	Query Match	90.2%;	Score 408.4;	DB 6;	Length 1535;	
	Best Local Similarity	97.8%;	Pred. No. 8.3e-56;			
	Matches 446;	Conservative	0;	Mismatches 6;	Indels	Gaps 3;

QY	1	gacgcggagcgcacatcaaaagtattcgcccttttcgccaccactgccccgcgcgg	60
DB			
DB	827	GATCGGCGAGCGCACGTCAAAGTGATTGGCGCCCTTTTCGCCCAACCTTGC	886
QY	61	tggatgtatcacccgcgcaggccgcgaagccgacactgcgcgaagaagcc--ctcaata	118
DB			
DB	887	TGGATGTCTCCACCCTCAGCGCGCGGAAGCCGACCTGCCGCGAAGCGCTCAATATC	946
QY	119	gtcccgcagtgctgcccgcgtacqccagcgggtcatggtgactggttacacccgcgcgg	178
DB			
DB	947	GTCCCGACGAGCTGCCCGCTAGCCGACGGGTATGGACTGGCTTACACCCGACGGCG	1006
QY	179	acctcaccgcacgcgaagcgcgcgcgaagcgc--gatcaccctgaacaccgcaatacg	237
DB			
DB	1007	ACCTACCGACACCGAACCGCGCCGAAACCGCGCATCACCTTGAGCAACCGACATACG	1066
QY	238	acggcatatcacgcgttaagtgtacctgacccccccaagtgcggggccacctttaaagcc	297
DB			
DB	1067	ACGGCATGTACCGGTAACTAGCTACCTGACCCCCCAAG-CGCGGGCCACCTTTGAAGCC	1125
QY	298	gtgctagcacaactggccgc	357
DB			
DB	1126	GTGCTAGCCAAACTGGCG	1185
QY	358	gacacaccccgatgcggc	417
DB			
DB	1186	GACACACCCCAGATGCGCGCGCGCGCATCGACGCGGACACCGCGCGCGCGCGCGCGCG	1245
QY	418	cacgacgggctgctggcgggctgctgctgctgctgctgctgctgctgctgctgctgctg	453
DB			
DB	1246	CACGACGGGTGCTGGCGGGGTGGCGCGGCTGATC	1281

RESULT	9
MTCY336/c	
LOCUS	MTCY336 32437 bp DNA BCT 03-AUG-2001
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.
ACCESSION	295586 AL123456
VERSION	295586.1 GI:3261785
KEYWORDS	.
SOURCE	Mycobacterium tuberculosis H37Rv.
ORGANISM	Mycobacterium tuberculosis H37Rv. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1. (bases 1 to 32437) Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E., Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squarres,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrell,B.G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence Nature. 393 (6685), 537-544 (1998)
JOURNAL	Nature. 393 (6685), 537-544 (1998)
MEDLINE	98295987
REFERENCE	2 (bases 1 to 32437)
AUTHORS	Parkhill,J.
TITLE	Direct Submission
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2117233.
COMMENT	

FEATURES	source
source	source
source	source
gene	gene
CDS	CDS

```
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAB09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:O06624"
/translacion="MKRSKMSGFAIGLAMLPVMAAPGLAAADPATRPVDYQOIID
VVIARGLSQRGVPSWAGGSGISGRTGTGINTVGEDASGLIOYVAGAGLKLPRSS
GOMYKGVKVLPOQARKDLIFYPGECTQSVALYLGKQMLEVGVQVQVSPVPTNGMT
PYLVRVLGTPQTPVQOAPVQAPVQOAPVQOAPVQOAPVQOAPVQOAPVQOAP
VQPPFGTARS"
complement(3030..3036)
/notes="possible RBS, AAGGAGG, for Rv1566c"
complement(3258..3263)
/notes="ASNI site: ATTAAT; probably linking fragments
B3/G6"
complement(3259..3543)
/gene="Rv1567c"
complement(3259..3543)
/gene="Rv1567c"
/notes="Rv1567c, (MTCY336.36), len: 94. Unknown. Tbpase
score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:O06623"
/translacion="NVTMTSWPSRLFAFTDNVCPDAPLVPPGVNVIYIPVMGGIG
AAIATAVTGPFPVSMKMGWMSFWPIISIAVITVTSIAGYAIAGFSERYWH"
3791..5104
/gene="bioA"
3791..5104
/gene="bioA"
/notes="Rv1568, (MTCY336.35c), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC2.6.1.62). O06622. Contains
aminotransferases class-III pyridoxal-phosphate attachment
site (P500600). FASTAresults, BioA MYCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1%identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
identity in 421 aa overlap. Tbpase score is 0.874"
/codon_start=1
/transl_table=11
/product="bioA"
/protein_id="CAB09069.1"
/db_xref="GI:2117269"
/db_xref="SWISS-PROT:O06622"
/translacion="MAAATGGLTPEQIIIVDGAHIMPIYSSIGREAVSPVVAVAIRGA
WLTIIRDQOPIEVLDDAMSSWMTAIGHGHGPAIDQALPTQLKVMNVMHFGGITHPEPAAR
LAKLLVDITPAGLDTVFSDGSYSVEVAAKMALQYWKRGILPKGKRRI.MTWRGGYHGD
TFLAMSICDPHGGHSLWTDVLAQVFAPOVPRDYDPAYSAAFEAOLAOHAGELAAVV
VEPVQAGGRRHFRDPRYLHLDRICIRYEVLLIFDELATGFGRTGALFAADHAGVSP
DIMCVKALTGGYLSLAATLCTADVAHTISAGAAGALMHGPTFWANPLACAVSVASVE
LLGGDWRTITELAAGLTAGLDRALPAVTDVRVCGATGVIECDRPVDLAVATPAA
LDRGVLRFPFRLNLYAMPPIICTPAEITQITSAMVEVARLVGSLP"
4541..4654
/gene="bioA"
/notes="PS00600 Aminotransferases class-III
pyridoxal-phosphate attachment site"
5101..6261
/gene="bioF"
5101..6261
/gene="bioF"
/notes="Rv1569, (MTCY336.34c), bioF, len: 386. Function:
probable 8-amino-7-oxononanoate synthase (EC 2.3.1.47)
O06621. Contains aminotransferases class-II
pyridoxal-phosphate attachment site (PS00599). FASTA
results, BioF MYCLE P4548(385 aa) opt: 1971; E(): 0;
80.1% identity in 381 aa overlap. Also similar to
MTCY1084.32, FASTA score: E(): 5.5e-29; 37.4% identity in
393 aa overlap
```

```
/codon_start=1
/transl_table=11
/product="bioF"
/protein_id="CAB09068.1"
/db_xref="GI:2117268"
/db_xref="SWISS-PROT:O06621"
/translacion="MKAATQARIDDSPLAWLDVAVQORHEAGRLRCLRPRAVATELD
LASNDYLGTSRIIPAVIDGGVQALRIWAGATGSRVLTGDTKLHQOFFEALAEFFVGAAA
GILFSSGYTANIGAVVGLSGPSLLIVSDAKSHASLVDAKRISARVAVVPIHIVDAVD
AALRSKQDRQAVVVTVDSVFSDGSLAPVRELEVCRRHGALLVDEAHGILGVKGGRG
LIYELCLGAPUVVMTTTLSSALSGOGVVLGPTPVRAHLIDAAKRP1FD7ICLAPAAV
GAARAAALVIOAEPMPQAVLNHAGELARMCGVAAVPSAMVSVILGEPESAVAAAA
CLDAGVKCCERPPTVPAGTSRLRLUTARASLNAGELELARRVLTIDVLAVKR"
5806..5835
/gene="bioF"
/notes="PS00599 Aminotransferases class-III
pyridoxal-phosphate attachment site"
6258..6938
/gene="bioB"
6258..6938
Query Match 90.2%; Score 408.4; DB 1; Length 32437;
Best Local Similarity 97.8%; Pred. No. 2.3e-56;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;
QY 1 galcggcgagggqacacatacaaaatgatlccgccccltttccacacctgccgcgcggcg 60
|||||
DB 17861 GATCGCGGAGGGCGACGTCAAAGTGTATTCGGTCCCTTTTTCGCCACCTGCCCGCCGCG 17802
QY 61 tggatgltgccarccgcgacgcccgaagcgcgaccctgccgcgaaggc--ctcaatc 118
|||||
DB 17801 TGGATGTGTCCACCCGCCAGCGCCGCCGAAGCCGACCTGCCCGGCAAGCCGCTCAATATC 17742
QY 119 gtccgcagagctgtgccgcgtacgcccagcgggtcatgactgggtacaccccgcgcgcg 178
|||||
DB 17741 GTCCCGACGAGTGTGCCCGCTACGCCACGCGGCTCATGACTGGCTACACCCCGACGCGC 17682
QY 179 acctcaccgacacgaaacgctgcgcgaacacgc-gcatcaccctgagcaccaccgaatcgc 237
|||||
DB 17681 ACCTCACCACACACGAACCGGCCCGCAACCGGGCATCACCTGAGCAACACGCAATACG 17622
QY 238 atgcacatgtcacgctgaaatgactacccacccacccacccacccacccacccacccaccc 297
|||||
DB 17621 AGCGCATGTCAAGGTAAAGTAACTAACTCAATCCCAAG-CCGCGGCCACCTTTCAGACC 17563
QY 298 gtgctagcacaactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 357
|||||
DB 17562 GTGCTAGCAACCTGGCTGCGTGGCGGCGGCGACCAACCGGACGACACACCCCGGTATC 17503
QY 458 gacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 417
DB 17502 GACACACATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 17443
QY 418 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
|||||
DB 17442 CACGACGGGCTGTGGCGCGGCTGGCGCGCTGGCGCGCTGATC 17407
RESULT 10
MTCY251/c
LOCUS 38380 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 5/162.
ACCESSION Z74410 AL123456
VERSION Z74410.1 GI:3261600
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 38380)
REFERENCE
AUTHORS
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
```

Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagers,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrell,B.G.

Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
 Nature, 393 (6685), 537-544 (1998)

98295987
 2 (bases 1 to 38380)
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1405752.

COMMENT

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source Location/Qualifiers
 1..38380
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 <1..35889
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y251"
 6..956
 /gene="hycD"
 6..956
 /gene="hycD"
 /note="RV0084, (MTCY251.02), len: 316, hycD, similar to HYCD_ECOLI P16430 formate hydrogenlyase subunit 4 (307 aa) FASTA scores, Opt: 570, z-score: 591.6, E(): 2.1e-26, (33.8% identity in 305 aa overlap) and NUO_H_ECOLI P33603nadh dehydrogenase i chain h (325 aa) FASTA scores, Opt: 207, z-score:220.7, E(): 9.5e-06, (26.5% identity in 260 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hycD"
 /protein_id="CAA98920.1"
 /db_xref="GI:1405754"
 /db_xref="SPTREMBL:Q10881"
 /translation="MSYLAQAQIGVVMGAPLIVIGMTROVRAWEGRAGAGLLOPWR DLLKQLGKQDITPAGTTIVFAAAPVIVAGTTLIAIAPLVATGSPLDPSADLFAVVG LLEFGTVLTLTGIDTGTGGMGASREITIAALVEPTILLAVEALSIPAGSANLGAL VASTIDHPGHVSVSLAGVAFVAVIVIVAEPTGLVDNPATHELETPMHEAVLETAG PRALVEMWAGMRLTVLLALLNLFLPWIAGNAPDLDVLTGGWAAVAAKVAIVLALL ATTEVLAKLRLFRVPELLAGSELLALLAVTAANFTTGA"

gene
 CDS
 967..1629
 /gene="hycP"
 967..1629
 /gene="hycP"

gene
 CDS
 1629..3095
 /gene="hycQ"
 1629..3095
 /gene="hycQ"

/note="RV0085, (MTCY251.03), len: 220, hycP, similar to SPLP7524|HYFE_ECOLI HYDROGENASE-4 COMPONENT E (216 aa) FASTA scores, Opt: 204 z-score: 272.1 E():1.2e-07; 25.5% identity in 216 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="hycP"
 /protein_id="CAA98921.1"
 /db_xref="GI:1405755"
 /db_xref="SWISS-PROT:Q10882"
 /translation="MSNANPSTLYDFDAAGGLVSLVLRWRDLRAIVRLAWQGAAL AAPULRGIRONDRAIAGVIAVLAALRALVLPWLLARVAGAEAAQRAETPLVNTASS LIIITAGLTITAFATPTQVNPVLEPGVTINAVPAFAVLIATFVMTTRLHVAYSOAGFL MDNGIAATAFLTAGVPLIVELGASLDLFAVIVIGVLTGRLRLRIFGDAADLKLREL RD"

gene
 CDS
 1629..3095
 /gene="hycQ"
 1629..3095
 /gene="hycQ"

/note="RV0086, (MTCY251.04), oxidoreductase, len: 488, hycQ, most similar to splp77437|HYFE_ECOLI HYDROGENASE-4 COMPONENT F(EC 1.- (526 aa), FASTA scores, Opt: 948 z-score: 1117.4 E(): 0; 35.9% identity in 493 aa overlap. Also similar to E. coli d908711 & NUO_L_ECOLI P33607 nadh dehydrogenase i chain 1 (613 aa) FASTA scores, Opt: 360, z-score: 354.9, E():3.2e-13, (27.9% identity in 488 aa overlap), and to NUO_H_ECOLI P33608, nadh dehydrogenase i chain h (425 aa), FASTA scores, Opt: 375, z-score: 371.4, E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hycQ"
 /protein_id="CAA98922.1"
 /db_xref="GI:1405756"
 /db_xref="SPTREMBL:Q10883"
 /translation="MTGLLLAAILAALAAASIASLITGWRTTATLTALSAATTVLACAV AMGFWMGSGAQFGILGLLRADALTVMVLVIGVIGVGLTAAASIGYDITELAHGHIDGR SAKLYGVLTPAFILCAMVLAVCANNGIVVIAIEATVITAFILVGHRTTLEATWKY VVICSGVIAVAFIGTGLVLYFAARDGAAAGALNLDLAEHAAGLDGFGVARLGLLL IGYGAKAGLFPFTWLDADHQAQAPVSAUMSGVLLAVAFSVLIRLPIDAVSGPAY LRNGLLVVGLATLILVAVMLITVDGVKRLAYSMMEHMGLITAAAGTTLAAIALL HVIANGIGKTVLPLAGLQJAAHDAITADITGMRRSRILGVSFAVGLIVLGLPPF AMPASLAIARSILARIALWGAALTAIATIGFTALARNSRMLLGTTPAAGAPATVP ATAAAALMVGILVVSAAALGITAGPLADLLGTAASNVGLP"

gene

CDS

3092..4570
 /gene="hycE"
 3092..4570
 /gene="hycE"

gene
 CDS
 3092..4570
 /gene="hycE"
 3092..4570
 /gene="hycE"

/note="RV0087, (MTCY251.05), oxidoreductase, len: 492, hycE, similar to HYCE_ECOLI P16431 formate hydrogenlyase subunit 5 (569 aa) FASTA scores, Opt: 680, z score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa overlap) and to NUO_H_ECOLI P33600 nadh dehydrogenase i chain d, (407 aa) FASTA scores, Opt: 245, z-score: 293.1, E(): 8.9e-10, (24.5% identity in 368aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hycE"
 /protein_id="CAA98923.1"
 /db_xref="GI:1405757"
 /db_xref="SPTREMBL:Q10884"
 /translation="MMSASLKHRSVERGLIATAEQWLADSFRLALVAHDDGDSLRV VTLFAGVPDRRVELEYVVPADNPEIRSLAYLFFPAGFEREMADLYGIRVGPDKPR RLVRHAHPDHPMRMTDAGPAFTDTGAFPLAVEGPGVYIIPVPGHAGLIEPGHFR FFSVAGETIVRLKARLWFRHGIKELFHRGPATAVDAERISGDTSAHAHLAISLAI EDALGIELPHVRLALIVELERLYNHAADGALANDVGYSLANAHOIRLENLLRR NAAVTHRLRLGAIAGGVALRALPDDELAALAVDLAEVATLTLANSQRIYDRKAGTA VLHPDASALGCLGYVASRGLRSRDARVEHTIPLPTEIGAPDGLIARVTVRDEFR AASALAQHIYESHTGPIEYAAALHPVGAPSSGIGVEGWRGTIVHRVEIDVDGRITR AKVDPSWFWNPALPVMADTIVDPFPLANKSFNQSYAGNDL"

gene
 CDS
 4605..5279
 /gene="RV0088"
 4605..5279
 /gene="RV0088"

[illegible]


```

78. .1481
/gene="Rv3448"
/notes="Rv3448, (MTCY77.20), len: 467 aa. Unknown membrane
protein, contains PS00402 Binding-comp signature. Some
transport systems inner membrane comp signature. Some
similarity to A021930|MTV035_18 Mycobacterium
tuberculosis (472 aa) fasta score, opt: 429 z-score:
488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3448"
/protein_id="CAB08736.1"
/db_xref="GI:3261767"
/db_xref="SPTREMBL:O33354"
/transl_table=11
/product="MPTSDGLRRVTHAGAAQVDLTLPAAPVPVATLIPSTVDILGR
CASPARATVOLSALGAPALPNATLTACQIRGQAVLVLHKSSAQPTPRCDQVAEAV
AAADTTARFQCCORTRLSGALAASCITAGGGLMLVRNALGTNVTNTRSDATAGVAAA
GLAALLFAVLCRTYRPIAGLTLSVITFAGVAGLLVPGVPGVHSLVAMAAAAA
TSVLAMITCGGITLFAVACCAVVAATLVGATAPVPAIGSLATLASFGLLVEVS
ARMAVLGASLRLPPALPDPDADLTDLTTRANKRADAWLTSLAAFAASATIGA
IGTAVATHGTHRSMGIGALAAVTGALLLRARSADTTRSLVFAICITTVATAFTVA
ADRALEHGPHTAALTAAVAAMFLGFAVAPALSVPVYRTIELLECLALIAMVPLTA
WLCGAYSARVHDLTWT"
misc_feature
611. .>43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y13E12"
1478. .2845
/gene="Rv3449"
1478. .2845
/gene="Rv3449"
/notes="Rv3449, (MTCY13E12.02), len: 455. Function:
probable precursor of serine protease. Has putative signal
peptideat N-terminus and hydrophobic stretch at
C-terminus. Contains three signatures typical of subtilase
family: aspartic acid active site (PS00136), histidine
active site (PS00137), serine active site (PS00138). FASTA
results: 053863 SERINEPROTEASE (390 aa) opt: 241; E():
8e-07; (38.0% identity in 387 aa overlap), similar to
MTCY15F10.29 (45.7% identity in 451 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3449"
/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/transl_table=11
/product="MPTSTRLLVLVSAALATLSGLTVAHAVSPPIIDERWLPSEL
PAPPRPVQREVCTEVTAEGRFRAERSAQLADLDQVRLTRGACURVAVIDTGA
RRRLRPVAGGDVFTGDTGCTADCAHGTIVAGLIAAPUAGLSNFSCVAPDVTLISI
ROSSKRPVDPSSSTGVGDVDTWAKAVKTAADLIGASVINISSIACVPAAPADRAL
GAALAYADVKNNAVIVAAAGNTGGAACPPAOPVTRDSVTVAVSPAWYDDYLVITGS
VNAQGPESAFITLAGPVDVDAATGEATSLSPFGDGTVNRLLGCGHGSIPISGTSYAPV
VSGLAALIRARFPLTARQVMORIESTAHHPAGMDPLVNGTVDALAAVSSDSIPQA
GTATSDPAPVAVPVRSTPGPSDRRALHTAFAGAACILALMATLATASRRRLPRGN
GIAGD"
misc_feature
1757. .1789
/gene="Rv3449"
/notes="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862. .1894
/gene="Rv3449"
/notes="PS00137 Serine proteases, subtilase family,
histidine active site"
2456. .2488
/gene="Rv3449"
/notes="PS00138 Serine proteases, subtilase family, serine
active site"
complement(2810. .4222)
/gene="Rv3450c"
complement(2810. .4222)
/gene="Rv3450c"
/notes="Rv3450c, (MTCY13E12.03c), len: 470. Function:

```

```

unknown but similar to Rv3895c MTCY15F10.17 (FASTA score:
1.6e-27; 36.4% identity in 475 aa overlap); possible
membrane spanning region near N-terminus, and to
Y14967|MLCB628_16 Mycobacterium leprae cosmid H628; (481
aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 32.9%
identity in 480 aa overlap, also similar to Rv3869
(MTV027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:O06317"
/transl_table=11
/product="MPSPATTLRVSGYRFLRRRIECALLFGDVCATGALRARTTSL
ALICLALIAVMGCAFAVAILRPOSALGOAPLVMGREGALYVVDVWHPVLNLASAR
LIAATANPQPVSESELGHTKRGPLLIGCAPOLLDOPLAGAESAWAICDSNCGSPT
VVGPAEDSSAONVITAEQMLLVATESGSPTYLLYGGRRVVDLADPAVVALRLOGRV
PVVAGSLLANVPEAPRITAPRIRGGRRASVGLPGVGVVTRIRASGDEYVYVLED
GVORIGVADLLRFGDSQGSVNVPTVAPDVIRVAPILVNTLPVSAPDRPPTVDDGP
GRAVTLICVITWTAQGAARVAFALGSPVPLGGVPTVLAQADGRGPDALDADVLPFG
RSAYVAARSLSGGTGTGYLVITDTGVRFATHDDDAHDLGLTAAIPAPVPVLATLPS
GPELSRANASVARIVAPGP"
complement(4049. .4081)
/gene="Rv3450c"
/notes="PS00137 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(4228. .4231)
/notes="possible RBS, GGAG, for Rv3450c"
4388. .5131
/gene="Rv3451"
4388. .5131
/gene="Rv3451"
/notes="Rv3451, (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MYCTU_Q50664
probable cutinase GY339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/db_xref="SWISS-PROT:O06318"
/transl_table=11
/product="MATAALITAVVLLIATGAVTPVAFAPGCDFAEFTFARGTGEPP
GLRVRGAPVNSLRQQLQMLIGVYVAVNAASRLQHGDDGANDAISHIKSMAGSCPT
KLVLAGYSGCATVITLVAGVPLGSLSEGSPLPAAYADNVAAVAFVGNPSKAGSLSS
LSPLGSKAIDLLNLTLPICLVGPRNEFSCHIDGYIPTTYTAAASFVQRLRAGSVPH
LHNSVPIHGSVIAQMPNCTAAVAPESLHGR"
4706. .4723
/gene="Rv3451"
/notes="PS00155 Cutinase, serine active site, GGYSQG"
5165. .5168
/notes="possible RBS, GAGG, for Rv3452"
5178. .5858
/gene="Rv3452"
5178. .5858
/gene="Rv3452"
/notes="Rv3452, (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,

```

```

Query Match 90.2%; Score 408.4; DB 1; Length 43401;
Best Local Similarity 97.8%; Pred. No. 2.1e-56;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

Qy 1 gatcgcgagcgccacatcaagtgatcgcgccttttgcgcacacctgccgcgcg 60
|||||
Db 15625 GATCGCGAGCGCAGCTCAAGTGATTCGCGCCCTTTTCGCCACCTGCCCGCGG 15684
|||||

Qy 61 tggatgtatccaccgcgcgagcgccgagcgacgacctgccgcgaaaggc--ctcaatatc 118
|||||

```


Db	15685	TGGATGTGTCACCCGCCAGCGCGGCGGAAGCCGACGACTGCCCGGCAAGCCCGCTCAATATC	15744
Qy	119	gtccgacgagctggcccgctacgcccagcgggctatgactgggtacacccccgacgcg	178
Db	15745	GTCCGACGAGCTGGCCCGCTACGCCACGCGGTATGACTGGCTACACCCCGACGGCG	15804
Qy	179	acctaccgacacccgacgcccgcgaacacgc-gataccctgagcaacacgaatacgc	237
Db	15805	ACCTACCGACACCGAACCGCGCGGCGGAACCGCGGCATCCCTGAGCAACCGCAATACG	15864
Qy	238	acgcatgtcacggctaaagtggctactgaccccccaagtgcgggcacacctttgaagcc	297
Db	15865	ACGCGATGTACCGCTAAGTGGCTACTGACCCCCCAAG-CGGGGCCACCTTTGAAGCC	15923
Qy	298	gtgtagcacaactggccgccccggcgacacccccgcagacaccccccggtcctc	357
Db	15924	GTGTAGCCAAACTGGCGCGCGCGGCGAGCAACCCCGACGACACACCCCGGTCAATC	15983
Qy	358	gacacacccccgatcgccgcgcacatcgaccgcgacacccgcagcacaagcccaacgcaac	417
Db	15984	GACACCAACCCCGATGGCGCGGCATCGACCGCGACACCCCGCAGCAACCCCAACGCAAC	16043
Qy	418	cacgacgggctgctggccggcgctgcgcgctgctc 453	
Db	16044	CACGACGGCTGCTGGCGCGGCTGCG-TCGCTGATC 16079	
RESULT	12		
LOCUS	AE007160	16384 bp DNA BCT 27-APR-2001	
DEFINITION	Mycobacterium tuberculosis CDC1551, section 246 of 280 of the complete genome.		
ACCESSION	AE007160	AE000515	
VERSION	AE007160.1	GI:13883389	
KEYWORDS	Mycobacterium tuberculosis CDC1551.		
SOURCE	Mycobacterium tuberculosis CDC1551		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
REFERENCE	1. (bases 1 to 16384)		
AUTHORS	Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.		
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains		
JOURNAL	Unpublished		
REFERENCE	2. (bases 1 to 16384)		
AUTHORS	Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1..16384		
	/organism="Mycobacterium tuberculosis CDC1551"		
	/strain="CDC1551"		
	/db_xref="taxon:83331"		
	/note="clinical strain"		
	96..839		
gene	/gene="MT3557"		
CDS	96..839		
	/gene="MT3557"		
	/note="similar to SP:Q00298; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="serine esterase, cutinase family"		
	complement(4323..4865)		
	gene		
	CDS		
	/protein_id="AAK47897.1"		
	/db_xref="GI:13883390"		
	/translation="MGAGALITAVVLLIALGAVWTLVAFADGCPDAEVTHARCTGEPP KILRGQAFVDSLRQOQTMEIGVYPVNYAASRLQOLHGGDGANDAI SHI KSMASCCPNT KILVGGYSGQATVIDI VAGVPLGS ISFGSPLPAAYADNVAAVFQNP SNKAGCSLSS I.SPLGSKAIDICNPDP ICHVGPNEFSGIDDI PTYTTOAASFVVORLRAAGSVPH I.PGSVPOLPGSVLQMPUTAAPAPESLHGR"		
	871..1566		
	/gene="MT3559"		
	871..1566		
	/gene="MT3559"		
	/note="similar to SP:Q00298; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="serine esterase, cutinase family"		
	/protein_id="AAK47898.1"		
	/db_xref="GI:13883391"		
	/translation="MREIGVTPRQPHSGRWAGARRLTSVAAAFAAALLLTLPAL APASAGCPDAEVVFAFGTGEPPGLGRVGOAFVSSLRQQTNKS IGTGYVAPANGDFL AAADGANDASDHIOQMASACRATRLVGGYSOGAAVTDIVTAAPLPGLGTQPLPAA IDHILAAIALFCNPSGRAGLMSALTPOGSKTINLCNNGDPICSDGNRWRAHLGYVPG MTNQANFVASKI"		
	1624..1863		
	/gene="MT3560"		
	1624..1863		
	/gene="MT3560"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK47899.1"		
	/db_xref="GI:13883392"		
	/translation="MTAADMHTAYHGLMLAPDPARQARRPSTOPNAHLSAKCAAMNRN IHTGREVCGHANRTLGI/TSVIAIALKKRYDACHHNQ"		
	1845..3523		
	/gene="MT3561"		
	1835..3523		
	/gene="MT3561"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK47900.1"		
	/db_xref="GI:13883393"		
	/translation="MMPGVITNSESPTAADHADRITATRETLEDYTLRLAPSKYHRWPP AVGGLSALGGLAYLADPAICANVGITWGTANALCGIAIEALVWFVTGLIPLAYYAARN IDHILITRSGSGGYGYSVTVNVEATFTFFFALESGTMAOGLKGLHPLIMWAGYAS ILITPLAVVYGMKVLSQLWTPLWLTLMAPFGYLVSHPHSIGUPFFSYAGKDGHG L.PGSVLLAAWVLELIALAQGLDYLLKPPHPTDENANHWMTLLAGKCVAFQA LKGLITGLPVLVYLMANTIGSSTLANQVPHQPMQITRTFVPCWMLATLAVILVLSQJK INVTNAYSCLAWINSFTRFKHYGRKVFAGVNLAIATLMEANMDFDINTILGCTA NTRMAWVAVASDIDCPNKYLLGLSKTPFPRGMILYAINPVGFGSLILAAGLSIVTFP GCGAALQPYSPILVAVITALVMPPLAAATKGYYLLRRTHLGLITLPMYDEHGNLSAAV I.PCHVCHDFRPMACOTHGAHVCSLCLSTDKOABHLVPLGLARAILPGIQVP"		
	complement(3485..4378)		
	/gene="MT3562"		
	complement(3485..4378)		
	/gene="MT3562"		
	/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U000096 PID:L788657; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="RNA pseudouridine synthase A"		
	/protein_id="AAK47901.1"		
	/db_xref="GI:13883394"		
	/translation="MSLTRRPKSPQRPPIRSGVYRLRLDIADYDGTDFAGWAAQVQG RTVAGLDLAALTTFRTVPLRAAGRTDAGVHAGQVAHVDPADALPNAYGVEPRAGHVG DPEFLPLRLRLGRELPAVRIIDITRAPAGDFARFSALRRHYVYRLSFAPYGVPEPOOA KYITAMPRELDLDANTASRLDGLHDFAEFCRHREGATTIRDLQRLDMSRAGTLVTA HVTDAFCWSVRSVLGALLAVGEHRRATTWCRELTLATGRSSDFAVAIPAIGLITLIQV DYPDDQLASRLNLYTRDRSG"		
	complement(4323..4865)		
	gene		

[illegible]

Search completed: April 2, 2002, 22:45:52

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 19:17:10 ; Search time 2977.34 seconds
(without alignments)
7718.491 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcgaa.....cgacgacaaacccgattgac 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htgo_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1393	100.0	43401	1	MTY13E12	295390 Mycobacteri
2	1354.2	97.2	3453	1	MTU13A340	U43540 Mycobacteri
C 3	1320.6	94.8	17783	1	AF007028	AE007028 Mycobacte
C 4	1315.8	94.5	10019	1	AF041819	AF041819 Mycobacte
C 5	1315	94.4	16384	1	AE007160	AE007160 Mycobacte
C 6	1314.2	94.3	1604	1	MBU35021	U35021 Mycobacteri
C 7	1307.4	93.9	32437	1	MTY3336	295586 Mycobacteri
C 8	1301.4	93.4	38380	1	MTY251	274410 Mycobacteri
C 9	1298.2	93.2	5764	1	AE006921	AE006921 Mycobacte
C 10	1281.8	92.0	9281	1	MBU43S1	U35017 Mycobacteri
C 11	1202.4	86.3	12412	6	186264	186264 Sequence 18
C 12	699.6	50.2	34331	1	MTCL165	295584 Mycobacteri
C 13	698	50.1	14953	1	AE006996	AE006996 Mycobacte
C 14	683.6	49.1	33100	1	MTY9F9	284498 Mycobacteri
C 15	682	49.0	14179	1	AE007053	AE007053 Mycobacte
C 16	675.6	48.5	40056	1	MSGY348	AD000702 Mycobacte
C 17	636.8	45.7	15019	1	AE007036	AE007036 Mycobacte
C 18	636.8	45.7	37432	1	MTCL125	298268 Mycobacteri
C 19	579.4	41.6	15504	1	AE006995	AE006995 Mycobacte
C 20	579.4	41.6	22550	1	MTY2208	295585 Mycobacteri
C 21	455	32.7	1535	6	AR096713	AR096713 Sequence
C 22	455	32.7	1535	6	AR147694	AR147694 Sequence
C 23	396	28.4	650	6	AR096715	AR096715 Sequence
C 24	396	28.4	650	6	AR147696	AR147696 Sequence
C 25	238	17.1	43254	1	MLCB1779	298271 Mycobacteri
C 26	238	17.1	312050	1	MLEPRTN3	AL583919 Mycobacte
C 27	212.8	15.3	348450	1	MLEPRTN1	AL583920 Mycobacte
C 28	137	9.8	36138	1	MLU15187	U15187 Mycobacteri
C 29	137	9.8	38109	1	MLCH5	295151 Mycobacteri
C 30	137	9.8	342300	1	MLEPRTN8	AL583924 Mycobacte
C 31	135.2	9.7	39193	1	U00021	U00021 Mycobacteri
C 32	135.2	9.7	332450	1	MLEPRTN5	296801 Mycobacteri
C 33	114.2	8.2	36225	1	MLCL581	AL583918 Mycobacte
C 34	114.2	8.2	344050	2	MLEPRTN2	Continuation (4 of
C 35	111.8	8.0	56870	2	CEY39B6_3	AC090654 Mus muscu
C 36	108.6	7.8	210827	2	AC090654	AC090654 Mus muscu
C 37	107.4	7.7	312050	2	MLEPRTN4	AL583919 Mycobacte
C 38	101.2	7.3	16387	1	AE006920	AE006920 Mycobacte
C 39	101.2	7.3	29256	1	MTV040	AL021428 Mycobacte
C 40	98.8	7.1	47852	2	MTV023	AL021428 Mycobacte
C 41	97.8	7.0	303091	2	AC084799	AC084799 Mus muscu
C 42	96.8	6.9	152547	2	ALI37857	ALI37857 Homo sapi
C 43	96.8	6.9	177883	2	AC046159	AC046159 Homo sapi
C 44	96.2	6.9	158167	2	AC044105	AC044105 Homo sapi
C 45	94.6	6.8	14472	1	AE003928	AE003928 Xylella f

ALIGNMENTS

RESULT	1
LOCUS	MTY13E12
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
ACCESSION	295390 AL123456
VERSION	295390.1 GI:3261766
KEYWORDS	
SOURCE	Mycobacterium tuberculosis H37Rv.
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium tuberculosis complex.
REFERENCE	1 (bases 1 to 43401)
AUTHORS	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,

Squares, S., Sqaures, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
 Nature. 393 (6685), 537-544 (1998)
 98295987
 2 (bases 1 to 43401)
 Parkhill, J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2104370.

COMMENT

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

Location/Qualifiers
 1. 43401
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 <1. 1481
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y77"
 78. 1481
 /gene="Rv3448"
 /note="Rv3448"
 78. 164
 /note="PS00402 Binding-protein-dependent transport systems:inner membrane comp signature"
 78. 1481
 /gene="Rv3448"
 /note="Rv3448, (MTCY77.20), len: 467 aa. Unknown membrane protein, contains PS00402 Binding-protein-dependent transport systems:inner membrane comp signature. Some similarity to AL021930|MTV035_18 Mycobacterium tuberculosis (472 aa) fasta score, opt: 429 z-score: 488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3448"
 /protein_id="CAB08736.1"
 /db_xref="GI:3261767"
 /db_xref="SPTREMBL:O33354"
 /translation="MPTSDGLRRVTTHAGAAVDLTLPAAVPVATLIPSTVDILGDR GASPTAARYOLSAIGAPALPNATLAQCGIRDCAGLVLLHKSSAQPTPRCDDVAEAV AALDTPAPQQRTRISGALAASCITAGGGLMLVRNALTNTNTRYSDATAGVVAAG LAALLFAVIACRYRDPDIAGLTSLVATIFGAVAGLLAVPGVGVHSLVAAAMAAAT TSLAMRTGGGTLTAVACCAVVAATLVGATAPVPAIGSLATLAFSGFLLEVS ARMAVLLAGLSPRLPGLNPDADALPTDLRTTRANRADWLTLAAAFATATTA IGTAVATGTHRRSGMGIALAAVTGALLLLRARADTRRSVFAICGTTTATFTAG ADRALEHCPWTAALTAMLAAMAVLGFVAPALSLSPVTYRTIELLECLLIAMVPLTA WLCGAYSAVRHLDLTWT"
 611. >43401
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"

source

gene
 CDS

/db_xref="taxon:83332"
 /clone="Y13E12"
 1478. 2845
 /gene="Rv3449"
 1478. 2845
 /gene="Rv3449"
 /note="Rv3449, (MTCY13E12.02), len: 455. Function: probable precursor of serine protease. Has putative signal peptideat N-terminus and hydrophobic stretch at C-terminus. Contains three signatures typical of subtilase family: aspartic acid active site (PS00136), histidine active site (PS00137), serine active site (PS00138). FASTA results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E(): 8e-07; (38.0% identity in 387 aa overlap), similar to MTCY15F10.24 (45.7% identity in 451 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3449"
 /protein_id="CAB08745.1"
 /db_xref="GI:2104372"
 /db_xref="SPTREMBL:O06316"
 /translation="MTTSKTLKLLVVSALATLSGLTPVAHAVSPPIDERWLPESAL PAPRPTVORECTEVTAEGRAGRAESAOALADLDVMRLTRGAGQAVVDTGVA RHRLKPVAGDYVFTGDTADCAHGLVAGIIAAPDAQSDNFSVAPDVLTISI RQSSSPAPVDPSSVSGVDVTMAKARVTAADLGASVINISIAICVPAARAPDDRAL GAALYAVDVKNVIVVAAAGNTGGAACQPPAGVTRDSVTVAVSPAWTDDUVLTVGS VNAQGPSTFTLAGPMDVAAATGEAVTSLSPFGDGTNRLGGHGSPIISGTSYAAAP VSCIAALIRARFPTLVARQVMQRIESTAHHPAGMDPLVGNCTVDALAAVSSDSIPOA GTATSDPAPVAVPVPRSTPGPSDRRLHTAFAGAAICLLALMATATASRRLRPGRN GIAGD"
 1757. 1789
 /gene="Rv3449"
 /note="PS00136 Serine proteases, subtilase family, aspartic acid active site"
 1862. 1894
 /gene="Rv3449"
 /note="PS00137 Serine proteases, subtilase family, histidine active site"
 2456. 2488
 /gene="Rv3449"
 /note="PS00138 Serine proteases, subtilase family, serine active site"
 /complement(2810. 4222)
 /gene="Rv3450c"
 /complement(2810. 4222)
 /note="Rv3450c, (MTCY13E12.03c), len: 470. Function: unknown but similar to Rv3895c MTCY15F10.17 (FASTA score: 1.6e-2; 46.4% identity in 475 aa overlap); possible membrane spanning region near N-terminus, and to Y14967|MLJ0628 16 Mycobacterium leprae cosmid H628; (481 aa) opt: 708.2 score: 710.9 E(): 4.5e-32; 32.9% identity in 480 aa overlap, also similar to Rv3869 (MTV027.04)"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3450c"
 /protein_id="CAB08716.1"
 /db_xref="GI:2104373"
 /db_xref="SPTREMBL:O06317"
 /translation="MPSPTTWLHVSGYRELLRRIECALLFGDVCAATGALRATTSL ALGCLAIIVAMGCAVALLRPQALQAPIVMGRESGALYVRVDVWHFVLNLASAR LIAATNAPQPVSESLGHTKRGPLIGIPGAPOLLQDQPLAGAESAWAICDSNCGSTT VVVGPAEDSSAQLVTABOIMLVATESGPTLYLLGGRRVVDLADPAVWALRQGRV PHVVAQSELLNAVPEAPRITAPRTGGGRASVGLPGVGGVVRITRASGDYVYVLED GVQIGVQADLRLFGDSQGSVNVPVADPVRVAPVINTLPSAFEDRPTPVGSGP GRVATLCVTWTTPAQCAARVAVLAGSPVPVGGVVDVTLAQADGRGALPDVLTLPGR RSVYVAARSLSGGGTGTRYLYLTDVTGVRFAIHDDDVADHLGLTAAIPAPMPVLTLPSP GPELSRANASVARDTVAPGP"
 complement(4049. 4081)
 /gene="Rv3450c"
 /note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

misc_feature

misc_feature

misc_feature

gene

CDS

misc_feature

RBS
gene
CDS

complement(4228..4231)
/note="possible RBS, GGAG, for Rv3450c"
4388..5131
/gene="Rv3451"
4388..5131
/gene="Rv3451"
/note="Rv3451, (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MVCTU_Q50664
probable cutinase cy339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/translation="MGAGALITAVVLLIALGAVTPVAFADGCPDAEVTFAFGTGEPP
GIRVGOAFVDSLRQTEIGVYVNTVAASRLQLHGGDGDANDAIHKSMASSCPNT
KIVLGYSGATVIDVAGVPLGSIFSGPLPAADVAAVAVFVGNPSNRAGSLSS
LSPFLGSAIDLCNFTDPICVHGPONEESHIDGVIPTVTTQAAEFVVRURAGSVPH
LPGSVPLGSLVQMPGTAAPAPESLHER"
4706..4723
/gene="Rv3451"
/note="PS00155 Cutinase, serine active site, GGYSGQ"
5165..5168
/note="possible RBS, GAGG, for Rv3452"
5178..5858
/gene="Rv3452"
5178..5858
/gene="Rv3452"
/note="Rv3452, (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,
Query Match 100.0%; Score 1393; DB 1; Length 43401;
Best Local Similarity 100.0%; Pred. No. 5.4e-181;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgggttcgagtgagccggaagatttcgaggtcttttgatgcgctggatgcccagctgg 60
Db 15252 TGGGTTCCGGGTAGCGCGAACGGATTCTCGAGGTCTTTGATGCGCTGGATGCCAGCTGG 15311
Qy 61 accgcttgagcagagtgctcttttgagtggttgaccaccccccagacggctgcggtctctgg 120
Db 15312 ACCGCTTGAGAGGTGCTCTTTGAGGTGTGTACACCCCGACAGACCGCTGCGTCTCTGG 15371
Qy 121 aacgtctggaatgcttggtgcgcgctaccgcggtggtgggtcagcgttgatcaaccaac 180
Db 15372 AACGCTTGGATGCTTGTGTGCGCGGTACCGCGGTGCGTACCGGTTCATCAACCAAC 15431
Qy 181 ttgacgcccagcagcagagaagaactggcggaacgctgtgtcgtgcgctggtggcaacc 240
Db 15432 TTGACGCCCAAGCCAGCGAGGAAGAACTGGGCGGCACGCTGTGCTGCGCGCTGCCAAC 15491
Qy 241 ggttacgcatcaccaagcccagaccccgccgcatcgccagcgcgcgcgcalctcgagac 300
Db 15492 GGTTACGCATCACCAAGCCCGACCGCGCGCGCGCATCGCGGACGCCCGCATCTCGAC 15551
Qy 301 ctctctgagcactcacgggtgaacgcttagcccccacagttgacggccacggccacccccc 360
Db 15552 CTCGTGAGCACTCACCGGTGACCGCTAGCCCCACAGTTGACCGCCACCGCCACCGCCC 15611
Qy 361 aacgcccaggcctgctcgagcgagcgcagctcaaaagtattcgccgcttttcgcccac 420
Db 15612 AACGCCAGGGCTGTATCGCGAGGCGACGTCAAAGTGTATTCGGCGCCTTTTCGCCCCAC 15671
Qy 421 ctgcccgcgagtgatgtgtccaccgcgagccgcgcgagcgcgacctggccgqcaaaa 480
Db 15672 CTGCCCCGCCGGTGGATGTGTCCACCCCGCAGCGCCGCGGACCTGGCGCGCAAA 15731

Qy 481 gccgtcaaatatcgctcccagcagagctggccgctaaqccccagcgggtctatggactggcta 540
Db 15732 GCGCTCAATATCGTCCGAGAGAGCTGGCCGCTACGCCAGCGGGTCAATGAGTGGCTA 15791
Qy 541 caccgcagcggcagactcaccgacacgaaacgagccgcccgaacacgagcgatcacccttgagc 600
Db 15792 CACCCGACGGGACCTCACGACACCAAGACGGCCGCCAACGGGCGATACACCTTGAGC 15851
Qy 601 aaccagcaalacgacgacatgtcacgctaaqggctacctgaccccccaagcgcgggccc 660
Db 15852 AACCAGCAATACGACGGCATGTACGGCTTAAGTGGCTACCTTGACCCCGCCCAAGCGCGGCC 15911
Qy 661 acctttgaagcctgtctagccaaactggcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 15912 ACCTTTGAAGCGCTGCTAGCAAACTGGCGCGCGCGCGCGCAACACCGCGACGACAC 15971
Qy 721 acccgtctcatcaacacaccccccgaatgagccacgacccacgacacccgcacaccaa 780
Db 15972 ACCCGGTTCACTGCAACACACCGCGCATGGCGCGCGCATCGACCGCGACCGCAGGCAA 16031
Qy 781 gcccaacgcaacacgacgagctgctgacgaggtcgccgctgctgacgctgctgctccgqgaaa 840
Db 16032 GCCCAACGCAACACGACGGGCTGCTGGTGGGCTGCGCGCTGATCGCTCCCGGAAA 16091
Qy 841 ctgggcccacacacacggtcttctccgctctgagtgtggtggtcaccacacacccctgacgcact 900
Db 16092 CTGGGCCCAACCAACAGGTTCTTCGGCTCTCGATCTGGTTCACCACACCGCTGACCGACCTG 16151
Qy 901 caaacgc 960
Db 16152 CAAACCG 16211
Qy 961 atccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020
Db 16212 ATCCCGCATGACCAACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16271
Qy 1021 attcttcacacgc 1080
Db 16272 ATCTTCGACCG 16331
Qy 1081 caagcagatcagctcttcac 1140
Db 16332 CAGCGGATCATGCTGTGTCGCCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16391
Qy 1141 gctacacacagcgaagc 1200
Db 16392 GCTATCACAGCGCAAGCCCGACCGCTACCGCTACCGCTACCGCTACCGCTACCGCTAC 16451
Qy 1201 accgacttaccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1260
Db 16452 ACCGAGCTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 16511
Qy 1261 cacaac 1320
Db 16512 CACAAACAAACCG 16571
Qy 1321 cccgcgcac 1380
Db 16572 CCGCGCGCAACACCTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16631
Qy 1381 aaacccgattgac 1393
Db 16632 AAACCCGATTGAC 16644

RESULT 2
MTU43540 3453 bp DNA BCT 14-AUG-1997
LOCUS Mycobacterium tuberculosis rfba, rhamnose biosynthesis protein
DEFINITION (rfba), and rmlC genes, complete cds.
ACCESSION U43540
VERSION U43540.1 GI:2326948
KEYWORDS

ORGANISM	Mycobacterium tuberculosis
REFERENCE	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
AUTHORS	Lee,T.
TITLE	Rhamnose biosynthetic genes related to a novel repeated sequence of Mycobacterium tuberculosis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3453)
AUTHORS	Lee,T.
TITLE	Direct Submission
JOURNAL	Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
REFERENCE	3 (bases 1 to 3453)
AUTHORS	Lee,T.
TITLE	Direct Submission
JOURNAL	Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
REMARK	Sequence update by submitter
COMMENT	On Aug 14, 1997 this sequence version replaced gi:1304703.
FEATURES	Location/Qualifiers
source	1..3453
gene	/organism="Mycobacterium tuberculosis"
gene	/db_xref="taxon:1773"
gene	/clone="p15116"
gene	150..1139
gene	/gene="rfba"
gene	150..1139
gene	/gene="rmb"
CDS	150..1139
CDS	/note="previously known as rfbB"
CDS	/codon_start=1
CDS	/transl_table=11
CDS	/protein_id="AAB66648.1"
CDS	/db_xref="GI:2326950"
CDS	/translation="MRLIVTGGAGFIGTNFVHSAVREHPDDAVTVLDALTYAGRRESLADVEDAIRLVGGDITDAELVSQVAESDAVVHFAESHVDNALDNPEPFLHTNIGTFVLEAVRRHGRVLRHHISTDEVYQLELDDRAFTSTPYNPSSPYSATKACADMLVRAWVSRICVRATISNCNNYPYQHVKEFTPTNCLTGLKLYGDDGNVRDWHVDDHNSAVRILDRIGIGETYLISSEGERDNLTVLSRLRLMDRPPDFDHWTDVRAHDLRYAIDPSRLDELCAWPKHDFERALKTTIDWTRDNESLWASLKHAGGRYQDAQ"
CDS	150..1139
CDS	/gene="rmlb"
CDS	/codon_start=1
CDS	/transl_table=11
CDS	/product="rhamnose biosynthesis protein"
CDS	/protein_id="AAB66647.1"
CDS	/db_xref="GI:2326949"
CDS	/translation="MRLIVTGGAGFIGTNFVHSAVREHPDDAVTVLDALTYAGRRESLADVEDAIRLVGGDITDAELVSQVAESDAVVHFAESHVDNALDNPEPFLHTNIGTFVLEAVRRHGRVLRHHISTDEVYQLELDDRAFTSTPYNPSSPYSATKACADMLVRAWVSRICVRATISNCNNYPYQHVKEFTPTNCLTGLKLYGDDGNVRDWHVDDHNSAVRILDRIGIGETYLISSEGERDNLTVLSRLRLMDRPPDFDHWTDVRAHDLRYAIDPSRLDELCAWPKHDFERALKTTIDWTRDNESLWASLKHAGGRYQDAQ"
CDS	1136..1747
CDS	/gene="rmlc"
CDS	1136..1747
CDS	/gene="rmlc"
CDS	/note="previously known as rfbC"
CDS	/codon_start=1
CDS	/transl_table=11
CDS	/protein_id="AAB66649.1"
CDS	/db_xref="GI:2326951"
CDS	/translation="MREPRELDVPGAMEITPTTHVDSRGLFFFWLTDHGFRAFAGHSLDVRQVNCSSVAGVLRGLHFAQLPPSOAKYVTCVSGSVDDVVVDIREGSPTFGRWDSVLLDDDRNIYVDSGLAHGLALQDNSTVMYLCASRYNPORHQHTICATDPTLASIRRWSDGCPSLSYNDXAAPSFEDVRKGLLPGGTIVSRF1GEMRGT"
CDS	2202..2655
CDS	repeat_region

BASE COUNT	679 a	1226 c	990 g	558 t
ORIGIN				
Query Match	97.28; Score 1454.2; DB 1; Length 3453;			
Best Local Similarity	99.28; Pred. No. 3.2e-175;			
Matches 1382; Conservative	0; Mismatches 8; Indels 3; Gaps 2;			
QY	1	199gttcggttaacgcgcuaacggatggtcgaagttctttgacgctggtatcccaagtg	60	
Db	1831	TGGGTTCGGGTAGCCGCCGACGATGTCGAGGTCTTGATCGCTGGATCCGAGCTGG	1890	
QY	61	accgcttgagcagtgctcttttgaagtggtgacacccagaaacggctgcggtctctgg	120	
Db	1891	ACCGCTTGACGAGGTGCTTTTGAGGTGTTACCAACCCACAGACGGCTGCGGTCTCTGG	1950	
QY	121	aacgctctgaatgcttggtgcgcgcgtacccgcggttgggtcaagcgttgatcaacaac	180	
Db	1951	AACGTCTGGAATGCTTGGT--GCCGGCTACCGCGGTGGGTACCGCTTGATCAACCAAC	2008	
QY	181	ttgacgcccacagccagcgagaaacactggcggcgacactgtgctgcgcgtggtgccaacc	240	
Db	2009	TGCACACCCAAAGCCAGCGAGGAAGAACTGGCGGGCACCCTGTGCTGGCGGTGGCCAACC	2068	
QY	241	ggtacgcatcaacaaagcccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	300	
Db	2069	GGTTACGCATCACCAAGCCCGACCGCGCCGCGCATCGCGCCGCGCATCTCGGAC	2128	
QY	301	ctcgtcgc	360	
Db	2129	CTCGTCGAGCAGCTACCGCGTGAACCCCTAGCCGCCACACATTGACCGCCACCGCCGCC	2188	
QY	361	aacgcacagcgtgacgc	420	
Db	2189	AACGCCACGGCTCATCGCGCAGCGCCACATCAAAGTCAATTCGCGCCCTTTTTCGCCAC	2248	
QY	421	ctgc	480	
Db	2249	CTGCCCGCGCGGTGGATGTGTCCACCGCCGAGGCGCGCGAAGCGAGCTGCGCCGGGAAA	2308	
QY	481	gcccctcaatatcgtccgc	540	
Db	2409	GCGTCAATATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2367	
QY	541	caccccgacgc	600	
Db	2368	CACCCGACGGCGAGCTCACCGCACACCGCCGCGCCGCAACCGCGCATCACCGTGAAC	2427	
QY	601	aaccagcaatcacgc	660	
Db	2428	AACCAAGCAATACGACGGCATGTACCGCTAAGTGGCTACCTGACCCCGCCGCGGCC	2487	
QY	661	accttgagcgcgtgtagccaaactgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	720	
Db	2488	ACCTTTGAAGCGTCTAGTACCAAACTGGCGCGCCCGCGCGGACCAACCCGACGACCCAC	2547	
QY	721	accccggtcatcgacacacaccccgatgcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	780	
Db	2548	ACCCCGG/CATCGACACACCCCGATCGCGCGGCATCGACCCGCGACACCCCGACGCCAA	2607	
QY	781	gcccaacgcacacgc	840	
Db	2608	GCCCAACGCACACGAGGGCTGCTGCGCGGGCTGCGCGCGCTGATCGCTCCCGGAAA	2667	
QY	841	ctggggcaacaacaggtctcttcgcgtctcgatcggtggtcaacacacacccctgacgcacgtg	900	

Db	2668	CTGGGCCAACACACGGCTCTTCCCGTCTCGATCGTGGTCAACACCCCTGACCGACCTG	2727
Qy	901	caaacccgcccggcaaggcttcaacggcgggcggaacccctgtaccacatgcccgtatgtg	960
Db	2728	CAAAACCGCGCGGCAAGAGGCTTACCGGGCGGGCACCCCTGTACCCATGGCGCATGTG	2787
Qy	961	atcgcgatgaccagccagccacccactactccccgcgaagcgggaggtaccccccagggc	1020
Db	2788	ATCCGATGACAGCCAGCCGCCACCTACTCTCCCGCAGCGGGAGGTACCCCGCGCG	2847
Qy	1021	atcttcgacacggcacacccccctggcgctgtatcacacccaaacgcctagcctccccggcc	1080
Db	2848	ATCTTCGACACAGCGCACACCCCTGGCGCTGTATCACACC-AAAGCGCTAGCCTCCCGGCC	2907
Qy	1081	cagcgatcatgtgttcgcaacgacggcgctgcacaaacccagcgtgaagcacccg	1140
Db	2908	CAGCGGATCATGTGTTCGCAACGACCGCGGTGCACCAACCCGGCTGTGACGCCCG	2967
Qy	1141	gcctaccacagccagccacccactcaccgctggaccagccagccgacgacccgacatc	1200
Db	2968	GCCTACCACAGCCAAAGCCACACCTGTCACCGCTGACCCAGCACCGACCGACATC	3027
Qy	1201	accgagctgacctgacctggcgcccgcaaacccgactcgccgaaagagctgagaccac	1260
Db	3028	ACCAGCTCACCTCGCCTGCGGCCCGCACCGGACTCGCGAAAGAGCTGACACACC	3087
Qy	1261	cacaacaacaccacggccacacccaatggtaccaccaccccccactcagccagcgcaa	1320
Db	3088	CACAACAACACCCACGCGCACACCGAATGGTACGACACCCACCCCTCCACACGCCCAA	3147
Qy	1321	ccccgacacaaccttcacaccccccgacgattcctccacaaccaagacgacgac	1380
Db	3148	CCCCGGCACCAACACCTTCACCAACCCCGAAGCATTCCTCCACAACCAAGACGACGAC	3207
Qy	1381	aaacccgattgac	1393
Db	3208	GAACCCGATTGAC	3220
RESULT	3		
AE007028/c			
LOCUS	AE007028	17783 bp	DNA
DEFINITION	Mycobacterium tuberculosis CDC1551, BCT section 114 of 280 of the complete genome.		
ACCESSION	AE007028	AE000516	
VERSION	AE007028.1	GI:13881250	
KEYWORDS	Mycobacterium tuberculosis CDC1551.		
SOURCE	Mycobacterium tuberculosis CDC1551.		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
REFERENCE	1 (bases 1 to 17783)		
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.		
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 17783)		
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1..17783		
	/organism="Mycobacterium tuberculosis CDC1551"		

gene	/strain="CDC1551"
CDS	/db_xref="taxon:83331"
	/note="clinical strain"
	complement(104..826)
	/gene="MT1622.1"
	complement(104..826)
	/gene="MT1622.1"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAK45890.1"
	/db_xref="GI:13881251"
	/translation="MIAKIAAPGATNPDDHTPV IDTTPDAAA I DRDTRSOAQRNHDGL IAGURALLASCKLGHNGHLPYSIVVTTLTDLQTGAGKGTGGTLLPMAUVIRMTSH AHHSPASGRYPQATIFDHGTGTPLALYHTRKRLASPAQRIMLFANDRGCTKPGCDAPAYHS QAHVHTASTGTGRTDITDLTLACDPDNRLEAKGWTHTKNTGHTEWLPPHLDHGQPR TMTFHHEKLLKHNDENHDPP"
gene	complement(831..1499)
CDS	/gene="MT1623"
	complement(831..1499)
	/gene="MT1623"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAK45891.1"
	/db_xref="GI:13881252"
	/translation="MIANSREELVEVEDALDAELDRLEDEVSFEVLTTPERLRSLERL ECLVRRLPAVGHITLNLDTQASBEELGTLCCALANLRITKPDAAIRLADADJGPR HALTCPELAPOLATATAOQGLGEAHKIVIRALFRPPARGGCVHPGCRKSKRGRO SRSLSSRRAGPLRPAGHGIATPRRRPHRHTRPQTRHHPEQPAIRRHVTAKWLPDPPS ANIL"
gene	1947..2996
CDS	/gene="MT1624"
	1947..2996
	/gene="MT1624"
	/note="similar to GP:1680650; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="biotin synthase"
	/protein_id="AAK45892.1"
	/db_xref="GI:13881253"
	/translation="MTQAVTKPTNADGQGGNSDILVVARQVLOHKGCLNODOVLA VLAHPHRLLELLALAEHVIMRWCGHFEVEVGLISLKTGGCPKCHPCSSGGLFASPV RSXWLLPISLVKAAKGLAKKATPRILVAVRGPDERIMAGVAGTETAIRNEVEINTA CGLDMTARAVVQVLAAGVVIKYNINILFTANSFFANVVTTHTEFRWVTLISNVRDAGME VTCGLTGMCTETLQRAETAAELALHAGPDEVPINFLNRPCTTFADI.FVMPVGDALKA VAAFRIALPRTMLRFAAGRCETLGDILGAKHGILCGINAVIVGNVLTTLGHPAEADLEL LHLQMPPLKALNASL"
gene	2997..3246
CDS	/gene="MT1625"
	2997..3246
	/gene="MT1625"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAK45893.1"
	/db_xref="GI:13881254"
	/translation="MVEIVAGKQRPVAAGVVNYTGELADTATPTAARMGLI:PPRRC AQCGRRMVVOVRPDGWMWRC:SRHGQVDSADLATQR"
gene	3317..3898
CDS	/gene="MT1626"
	3317..3898
	/gene="MT1626"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAK45894.1"
	/db_xref="GI:13881255"

Db 6280 GTCATGACACACCCCGATGCGCGCATCGACCGGACACCGCGACGCCAA 6221
 Qy 787 CGAACACAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 846
 Db 6220 CGAACACAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 6161
 Qy 847 CAACACAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 906
 Db 6160 CAACACAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 6101
 Qy 907 GCGCGGCAAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 966
 Db 6100 GCGCGGCAAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 6041
 Qy 967 ATGACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1026
 Db 6040 ATGACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5981
 Qy 1027 GACACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1086
 Db 5980 GACACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5921
 Qy 1087 ATGACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1146
 Db 5920 ATGACAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5861
 Qy 1147 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1206
 Db 5860 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5801
 Qy 1207 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1266
 Db 5800 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5741
 Qy 1267 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1326
 Db 5740 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5681
 Qy 1327 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 1386
 Db 5680 CACAGAGAGAGGGGTGTCGCGGGGTGCGCGGCTGATCGCTCCGGGAACGGGC 5621
 Qy 1387 gat 1389
 Db 5620 gat 5618
 RESULT 5
 AEO07160
 LOCUS 16384 bp DNA BCT 27-APR-2001
 DEFINITION Mycobacterium tuberculosis CDC1551, section 246 of the complete genome.
 ACCESSION AEO07160 AE000516
 VERSION AEO07160.1 GI:13883389
 KEYWORDS
 SOURCE Mycobacterium tuberculosis CDC1551.
 ORGANISM Mycobacterium tuberculosis CDC1551.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 16384)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Mikula, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
 Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
 Unpublished
 2 (bases 1 to 16384)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,

Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Deicher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
 Direct Submission
 Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..16384
 /organism="Mycobacterium tuberculosis CDC1551"
 /strain="CDC1551"
 /db_xref="taxon:8331"
 /note="clinical strain"
 96..839
 /gene="MT3557"
 96..839
 /gene="MT3557"
 /note="similar to SP:Q00298; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="serine esterase, cutinase family"
 /protein_id="AAK47897.1"
 /translation="MGAGALITAVVLLIALGAVTVLFAADGCPDAEVTFAAGTGEPPGIGRVGAFVDSLRQQTGMETIGVVPVNAASRLQHGCDGANDAISHIKSMASCPNTKLVLGGYSOGATVTDIVAGVPLGSIISFGSLPRAADVAVAVFGNPSNAGGSLSSLPVFGSKAIDLCNPTDPICHVGPNGNEFSGHIDYIFPTTQAASFVVQIKAGSVPHLPQSVPLPGSVLQMPTAAPAPESLHGR"
 871..1566
 /gene="MT3559"
 871..1566
 /gene="MT3559"
 /note="similar to SP:Q00298; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="serine esterase, cutinase family"
 /protein_id="AAK47898.1"
 /translation="MRFGVIPRPHSGRWAGAAARLTSLVAAAFATLITLPAAPASAGCPDAEVVFAFGTCPEPGICRGVQAFVSSLRQTNKSGITGVNYPANGDFLAAAGDASDHILQMASACRATRLVLGGYSQGAVIDVTAAPPLGLGPTQLPPAAADHIAIALFGNPSGRAGGLMSALTPTQFGSKTINLNCNGDPTCDGNRRWRAHLGVYPMTNOAARFVASRI"
 1624..1863
 /gene="MT3560"
 1624..1863
 /gene="MT3560"
 /note="identified by Glimmer2: putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK47899.1"
 /db_xref="GI:13883392"
 /translation="MTAADWHTAYHGLMLAPDPAQARRPSTQPNNAHLSAECAAMRNHFTGREVCGHANRTLGLTATATLRVDAGRHKKQ"
 1835..3523
 /gene="MT3561"
 1835..3523
 /gene="MT3561"
 /note="identified by Glimmer2: putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK47900.1"
 /db_xref="GI:13883393"
 /translation="MMPGVITNSPSTADHDRTATRTLEDYTLRLAPRSYRRMPPAVVSAUGGIAYLADFAGVNGVITGWGTANALCGIAPLVVFTGLPLAYAYARNIDLDLITSGSGFYGVSVTVTFATFTFFALEGSIQAQGLKGLHILVWAGYACSTLIIFPLVYVGNKVLSQLQMTWTLPLILMAAPFGLVYVSHDPSTGQFYSYAGKDHGGLSFGSVLLAAGVCLSLTAQIAEQIDYLRFPMPPTPENANRWMTTLAAGPCWAFGA TKQIIGLFLAVYLMANIPGSSSTIANOPVHQMIVTFTVPGHALTLAVILVLSQIK INVTNAYSGLAWTNSFTRLTKHYFGRVVFLGVNLAILMEANFDLNTILGCTA

Db	11499	CACCCGACGGGACGACCGGACCGGCGCCGCAACGGGGCATCACCTGAGC		11558
Qy	601	aaccagcaatcacgacgcatgtcaacggctaaagtgtactacaccccccaacgagcgagcc		660
Db	11559	AACGAGCACTACGACGATGTCTACGGCTAAGTGGCTACTGACGCCCCAAGCGGGGGCC		11618
Qy	661	acctttgaacccgtgtctagccaacttggccgccccggcgcgaccacacccccgacgacac		720
Db	11619	ACCTTTGAAGCCGTGTAGCCAACTGGCGGCCCCCGGGCGGACCAACCCGACGACCAAC		11678
Qy	721	accccggtcatcgacacaccccccgatgcyggccgcacatcgacgcgagacacccgacgaca		780
Db	11679	ACCCGGTCTATCGACACACCCCGATGCGCGCGCGCATGACCGCGACACCCGACGCCAA		11738
Qy	781	gcccaacgcaaacacgacgagctgctgcyggcggtctgcgcygctgctgctctccgggaa		840
Db	11739	GCCCCAAGCGCAACACGAGGGGTCTGGCGCGGGCTGCACGGCTGATCGCTCCGGGAA		11798
Qy	841	ctggggcaacacacggtcttcccgctcgtatggtgaccacacccctgacgcgacctg		900
Db	11799	CTGGGCGCAACACAACGGGTCTTCCCGTCTCGATGTGGTCAACCAACACCTGACCGACTG		11858
Qy	901	caaacggcgccgcaacgaggtctaacggcgcgcgacacctgtaccctatggccatgclg		960
Db	11859	CAAAACGGCGCGGCAAGGGGCTTACCAGGGCGGGCACCTGTACTACCATGGCCGATGTG		11918
Qy	961	atccgcgatgacagccacgcccacactactctcccgcaagcggggaggtaccctcccgag		1020
Db	11919	ATCCGATGACCCAGCCAGCGCCACCACTACTTCCCGCCGAAGCGGAGGTACCCCGAGCG		11978
Qy	1021	attctcgaccacggcaacacgctgagcgctgtatcacacaaaacgctagctcccgccc		1080
Db	11979	ATCTTCGACCAAGCGCACACCCCTGGCGCTGTATCACACCAAGCGCTTAGCCTCCCGGCC		12038
Qy	1081	caagcgatcatgctgttcgcaacgacgcgcggtgcaccaaaccggcgctgtgacgcacg		1140
Db	12039	GACCGGATCATGCTGTTCGCCACAGCGCGGGCTGCACCAACCCGGCTGTGTGACGCACCG		12098
Qy	1141	gcttaccacagccaaagccacacacgctacgcgcttgaccagcaccgagacacacgacatc		1200
Db	12099	GGCTATACACAGCAAGCCACACAGCTACCGGCTTGGACCAAGCACCGGACGACCGAGTC		12158
Qy	1201	accgagctgaccttgcttcgccccgacaaacgacactcgcgaaaaaggtgtagccacc		1260
Db	12159	ACCGAGCTGACCTTGCTCGCGGCCCGACAAACCGACTTCCCGGAAAAAGGTGACCAACC		12218
Qy	1261	cacaaacacacacgccccacacgcatggtctacacacacccactctgagacacgctaa		1320
Db	12219	CACAAAAACACCCAGCGGCACACCGGATGGGTACACCAATCCACCTCGACGAGGCTAA		12278
Qy	1321	ccccgcacacacacttccacacccc		1347
Db	12279	CCG"TGACCTGTGATACATACATACAC		12305
RESULT	6			
MBU35021	1604 bp	DNA	BCI	27-APR-1996
LOCUS	Mycobacterium bovis	DNA flanking deletion region 3.		
DEFINITION	U35021			
ACCESSION	U35021.1	GI:1049243		
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				


```
787 cgcaacacagcggctgctgccgggctgcgcgcgctgacgcctccgggaactgggc 846
Db 864 CGCAACACAGCGGGTGTGTCGGGGCTGCGGGCTGATCGCTCCGGGAAGTGGGC 923
Qy 847 caacaaagctgttcccgctcgcgtcgtggtcacaacacccctgaccgacctgcaaac 906
Db 924 CAACACAAGGCTTCCCGCTCGATCGTGGTCACACACCCACCCCTGACCGACTGCAAAAC 983
Qy 907 ggcgcgcgaagggcttcacggcgccgacccctgctaccatgacgcgagtgtgacgcg 966
Db 984 GGGCGCGGCAAGGGCTTCACAGCGCGCGGACCCCTGCTACCCATGCGCGATGTGATCCGC 1043
Qy 967 ataccagcaccacccactactccccgcgaagcggaggtaccctcccgagcgatcttc 1026
Db 1044 ATGACCAAGCAGCCACCACTACTCCCGCGCAAGCGGAGGTACCCCGCAGCGATCTTC 1103
Qy 1027 gaccagggcacacccctggcgtgtgtatcacacaaacgctagctcccgcccgccagcg 1086
Db 1104 GACCAGGCAACACCCCTGGCGCTGTATCACACCAACGGCTAGCCCTCCCGGCCACGCG 1163
Qy 1087 atctgctgttcgcaacgacggcgtgcaccaaaccggctgtgacgacccgacctac 1146
Db 1164 ATCATGCTGTTCGCAACGACCGGGCTGCACCAACCCGGTTGTGACGACCCGGCTAC 1223
Qy 1147 cacagcgaagccaccagtcacgcctgacgcagcagcagcagcagcagcagcagcag 1206
Db 1224 CACAGCAAGCCACACCGTACCGGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1283
Qy 1207 ctgacctggcctggcccccgaacacccgactcgccgaagaaagctggaccacccacaac 1266
Db 1284 CTCACCTGGCGCTCGACACCGCGGCTGCACCAACCCGGTTGTGACGACCCGGCTAC 1343
Qy 1267 aacaccacggccacacgaatggctaccacccacccacccacccacccacccacccgc 1326
Db 1344 AACACCCAGCCACACCGAATGGCTACCAACACCCACCCACCCACCCACCCACCCGC 1403
Qy 1327 accaaccttcaccccccgaacgattctccacacaaacgacacgacgacgacaccc 1386
Db 1404 ACCAACACTTCCACCAACGAGAGCTGCTAGCCCAACACGACGAGGACACACCGAC 1463
Qy 1387 gat 1389
Db 1464 GAT 1466

RESULT 7
LOCUS MTCY336 32437 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; sequenced 70/162.
ACCESSION Z95586 AL123456
VERSION Z95586.1 GI:3261785
KEYWORDS
SOURCE
ORGANISM
  Mycobacterium tuberculosis H37Rv.
  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Corynebacterineae; Mycobacteriaceae;
  Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
  1 (bases 1 to 32437)
  Parkhill, J., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
  Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
  Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
  Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
  Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
  Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
  Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellon, S.,
  Squares, S., Squires, R., Stulson, J.E., Taylor, K., Whitehead, S. and
  Barrall, B.G.
  Deciphering the biology of Mycobacterium tuberculosis from the
  complete genome sequence
  Nature 393 (6685), 537-544 (1998)
JOURNAL
MEDLINE 98295987
REFERENCE 2 (bases 1 to 32437)
```

AUTHORS
TITLE
JOURNAL

Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2117233.

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

Location/Qualifiers
1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
c1..273
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
c3..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
cclone="Y48"
complement(39..2228)
/gene="Rv1565c"
complement(39..2228)
/gene="Rv1565c"
/note="Rv1565c, (MTCY336.38), len: 729. Unknown membrane protein, some similarity to O05402 HYPOTHETICAL 72.2 KD PROTEIN from H. subtilis (634 aa) opt: 384 E(): 4.8e-17; 29.1% identity in 478 aa overlap and to hypothetical protein in H. influenzae. N-terminal half hydrophobic.
FASTA results: Y492 HAEIN P4393 hi0392 (245 aa) opt: 265; E(): 5.5e-10; 28.1% identity in 247 aa overlap. TUparse score is 0.940
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1565c"
/protein_id="CAB09072.1"
/db_xref="GI:2117272"
/db_xref="SPTSRMBL:006625"
/translation="MLTUSPPRPALPEPALPVTMGTTRTGTGYRHDLDLRCVATA LVAVFWFVRVSGVDYFLALSGFFGGKILRAALNDLSLAEIVIRILRLLEPA LVVILAGCALLTIAIQTPTRWFAFANQSLAGYQNMELASTVSNYLRAEAFVPIQ HWSVVGQFYLAFLLLVAGCAYLLRRLFRGRAPYLRMTFVLLSTLTLASTIYAI VAHAYQATVYNTFAREWELLAGAVVPHVPMWLTAVATAAIIISGAL IDGVEFFGPNALVPVGNATMLILAGANRGHPCGTDRDLPLNRLATAPALVALCAMA RSPAAPPPIPNRLRLRPTI VIGSVVALLGVALTATSTFWREHVIVQRAKELSGLS YSWLWHPPLLIIFWLSYTGHRHANFVEGAALLVSGLLAYLTTLVEDPLRYAPAGV SDPYPCRALIDHVRPKLRMRPTVLEVRHDLPTSKDGCISDFNPALJNCTTGDDV IPTTALAGGSHAEHLTALDILGRMHFKVVTYLMKGCPLSTEVEPLIMGNAPYPO CHQWQAAAKLVADHPDYVFTTSTPRNNIKPGDVPMTYIQWTFADNNIPVLMR DTPWLYKDGOPPIPADCLAKGNQSCGIAKSKVLVDNRNPTLDFVAREPLKPLDMSD AICRTDTCRAVEGNVLVRDLSHLTPTTYRTMTSELGRIANTDWW"
complement(2327..3019)
/gene="Rv1566c"
complement(2327..3019)
CDS

```

/ gene="Rv1566c"
/ note="Rv1566c, (MTCY336.37), len: 230. Function: unknown
probably exported has QQAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasin
1.2(007390, 007391) from M. avium and slightly similar to
C-terminus of Listeria invasion-associated protein p60
precursor p60.LiSMO P21171. FASTA results, Q49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. TParse score is 0.890"
/ codon_start=1
/ trans1_table=11
/ product="hypothetical protein Rv1566c"
/ protein_id="CAB09071.1"
/ db_xref="GI:2117271"
/ db_xref="SPTREMBL:O06624"
/ translation="MKRSKSGSFAIGLAMLPWVAPGLAAADPATRPVDYQQIITD
VVIAGLSORGVFSSWAGGSGSPTRGTGINTVFDSAGLIQYAVAGAGLKIPRSS
GOMYKGVKQVLPQARKGDLIFYGPEGTSVALYLGKQMLEVDGVQVSFVRNGMT
PLYVLVGTQPTPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAP
VQPPPGTGARSK"
complement(3030, 3036)
/ note="possible RBS, AAGAGG, for Rv1566c"
complement(3258, .3263)
/ note="ASNI site: ATTAAT; probably linking fragments
B3/66"
complement(3259, .3543)
/ gene="Rv1567c"
complement(3259, .3543)
/ gene="Rv1567c"
/ note="Rv1567c, (MTCY336.36), len: 94. Unknown. TParse
score is 0.915"
/ codon_start=1
/ trans1_table=11
/ product="hypothetical protein Rv1567c"
/ protein_id="CAB09070.1"
/ db_xref="GI:2117270"
/ db_xref="SPTREMBL:O06623"
/ translation="MVTMTSNPSRLFAFTDNCPPDAPLVFPGVNYIYIPVMWGGT
AAIATAVIGPVSMLKGMYSFWPIISIAVITVTSIAGYAIAGFSRYNH"
3791, .5104
/ gene="bioA"
3791, .5104
/ gene="bioA"
/ note="Rv1568, (MTCY336.35c), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC2.6.1.62) O06622. Contains
aminotransferases Class-III pyridoxal-phosphate attachment
site (PS00600). FASTA results, BIOA_MVCLP P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E(): 3.5e-16; 29.5%
identity in 421 aa overlap. TParse score is 0.874"
/ codon_start=1
/ trans1_table=11
/ product="bioA"
/ protein_id="CAB09069.1"
/ db_xref="GI:2117269"
/ db_xref="SWISS-PROT:O06622"
/ translation="MAAATCGLTPEQIIAIVDGAHLWHPSYSSIGREAVSPVVAAGA
WFLIIRDGPQIEVLDAWSSMTAIHGHGHPALDQALTOLRVNHNHMEGLTHEPAR
LAKLLVDITPAGLDVFFSDGSVSVEVAAKNALQYWRGRGLPGRLMTWRGGYHGD
TFPLMSICDPHGSMHSLDWDVLAQVAFQVPRDYPAYSAFAEQALQAHGELAAYV
VEPVQAGAGMRHDPRLYLHDLRCIRRYEVLIIIDEIATGFGRTGALFAADHAGVSP
DINCVGALTGTLSUATLCTADVAHTISAGAAGALMHGPTFMANLACAVSVASVE
LLIGDQMRTRI TELAAFLPGLDTRALPVDVRCVGAIGVIECDRPVDLAVATPAA
LDGWVLRFRNLVNLAMPYICTPAEITITISAMVEVARLVGSLP"
4541, .4654
/ gene="bioA"
/ note="PS00600. Aminotransferases class-III
pyridoxal-phosphate attachment site"
5101, .6261
/ gene="bioA"
5101, .6261

```

```

/ gene="bioF"
/ note="Rv1569, (MTCY336.34c), bioF, len: 386. Function:
probable 8-amino-7-oxononanoate synthase (EC 2.3.1.47)
O06621. Contains aminotransferases class-II
pyridoxal-phosphate attachment site (PS00599). FASTA
results, BIOF_MVCLP P45487(385 aa) opt: 1971; E(): 0;
80.1% identity in 381 aa overlap. Also similar to
MTCY10H4.32, FASTA score: E(): 5.5e-29; 37.4% identity in
393 aa overlap"
/ codon_start=1
/ trans1_table=11
/ product="bioF"
/ protein_id="CAB09068.1"
/ db_xref="GI:2117258"
/ db_xref="SWISS-PROT:O06621"
/ translation="MKAATQARIDDSPLAWLDVAVQORHEAGRLRCLRPAPVATELD
IASNDVYLGLSRHFAVINDGVOALRIMGATGASRLVTGDTKLHQFEAEAEFVGAA
GLLFSSGYTANLGAIVGSLGPGSLVSDRASLASLDACRLSRVYVTVPHRDVDAVD
AALSRDEQRADVVTDSVFSADGSLAPVRELLEVCRRHGLALLDEAHIGLVGGGRG
LLYLEGLGAPDVVMTTLLSKALSGQGVVVGPTVPVRAHLIDAAEPFI DTGLAPAAV
CAARAALRVIQAEPRPQAVLNHAGELARMCGVAAVPDSAMYSVLLGPESAVAAAA
CLIDAGVKVGCFRPPTVPAGTSRLRLTARASLNAGLELARRVLTDLVLAARR"
5806, .5835
/ gene="bioF"
/ note="PS00599. Aminotransferases class-II
pyridoxal-phosphate attachment site"
6258, .6938
/ gene="bioD"
6258, .6938

```

	Query Match	93.9%;	Score 1307.4;	DB 1;	Length 32437;
	Best Local Similarity	98.4%;	Pred. No. 2.5e-169;		
	Matches 1320;	Conservative	0;	Mismatches 21;	Indels 0;
	Gaps	0;			
QY	7	cgggtagccgcgaacggtgttcgaggtcttttgcgtcgctggatgcgagctgacgcgt	66		
Db	18228	CGAATAGCCGCGGAGGACCTGTGCGAGTCTTTGATCGCTGGATGCGAGCTGGACCGT	18169		
QY	67	tgacagaggtgtcttttgaggtgtttgacacacccacgaacgagctgcggtctctggaacg	126		
Db	18168	TGGACGAGGTGTCTTTTGAGGTGTTGACCAACCCGCGGCTGCGGTCTCTTGAACGTC	18109		
QY	127	tggaaatgcttggcgcgcgcgtlaccgcggtgaggtcacgctgtgatacaaccaacttgacg	186		
Db	18108	TGGAATGCTTGGTTCGCCCGGCTACCCGCGGTGGGTGCGGTGATGATCAACCACTTGACG	18049		
QY	187	cccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	246		
Db	18048	CTCAAGCTAGCTAGCAAGAAATTTGAAATGCTGCTGCTGCGCGCTGGCCACACCGGTTAC	17989		
QY	247	gcatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	306		
Db	17988	GCATACCAAGCCCGACCGCCCGCGCGCATTCGCGACCGCGCGCATTCGCGACCTCGTC	17929		
QY	307	gagcactcaccggtgaacgcgtlagccccacagttgacgcgacccgacccgacccgacccg	366		
Db	17928	GAGCACTACCGGTGAACCGCTAGCCCCACACTTACCGCCACCGCCGCGCCCAACGCC	17869		
QY	367	agggcctgctgcgagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	426		
Db	17868	AGGGCTGATGTCGCGGAGGCGCACGTCAAAGTGATTTCGCGCCCTTTTTCGCCACCTGCC	17809		
QY	427	gcccgcggtgagtggtccaccccgccagcccgccgagcagcagcagcagcagcagcagcagcagc	486		
Db	17808	GCCGCGGTGGATGTCTCACCCCGCGCGCGCATTCGCGACCGCGCATTCGCGACCTCGTC	17749		
QY	487	caatcgtcccgacgagcgtggccgcgtacgccccagcgggtcatgagctggctacacccc	546		
Db	17748	CAATATCGTCCCGACGAGCTGGCCCGCTACGCCAGCGGTCATGGAGCTGGCTACACCCC	17689		
QY	547	aacgcgacctcaccgacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	606		
Db	17688	GACGCGACCTCACCGACACCGACGCGCGCGCGCAACCGCGGCTCATCCCTGAGCAACCG	17629		

QY 607 caatcagcgcatgtcacgctgaagtgcctacctgaccccccaagcgcgggccaactttt 666
|||||
Db 17628 CAATACGAGCGCATGTACGGCTAAGTGGCTACCTGACCCCCCAAGCGCGGGCCACCTTT 17569
QY 667 gaagccgtgtcagcaactggccgccccccggcgcgagcaaacccccgcagcaccaccccg 726
|||||
Db 17568 GAAGCCGTGTAGCAAACTGGCGCGCCCGCGCGCGACCAACCCCGACGACACCCCG 17509
QY 727 gtcacacacacccccccgcatggcgccgcatcagcagcgagaccccgaccccgagccca 786
|||||
Db 17508 GTCATCGACACACCCCGCATGCGCGCGCCATCGACCGACACCCCGCAAGCCCAA 17449
QY 787 cgcacacacagcggtcgtggcgggcggtcgcgcgctgctgcgctcccgagaaactgggc 846
|||||
Db 17448 CGCAACACGAGCGGTGCTGGCGGGCTGGCGCGCTGATCGCTCCGGGAACATGGCG 17389
QY 847 caacacacgggtcttcctcgtctcgtatcgtgtgttcacacacccctgacccgacctgcaaac 906
|||||
Db 17388 CAACACACGGGTCTTCCCGTCTCGATGTGTTCACCAACACCCCTGACCGACCTGCAAAAC 17329
QY 907 ggcgcgcggaagggttcacccgcgcgccgacccctgtaccatgcccgtatgacccgc 966
|||||
Db 17328 GCGCGCGCAAGGGTTCACCGCGCGGGGACCCCTGTACCATGGCGGTGTGATCCCG 17269
QY 967 atgacgcgcacgccccaccactactcccccgcgagcgagggtacccccaggcgatcttc 1026
|||||
Db 17268 ATGACGACGCGCCACCACTACTCCCGCAAGCGGAGGTACCCCGACGGCGATCTTC 17209
QY 1027 gaccacgcacacccctggcgctgtatcacacaaagcctagcctccccggccagcg 1086
|||||
Db 17208 GACCACGGCACCCCTGGCGCTGTATCACCAAAAGCGCTAGCCCTCCCGGCGCCAGCG 17149
QY 1087 atcatgtgttcgcaacgacggcggtgcacacaaacgggtgtgacgacccgacctac 1146
|||||
Db 17148 ATCATGTGTTCGCAACGACCGCGCTGCACCAAAACCGGCTGTGACGACCGGCTTAC 17089
QY 1147 cacaagcaacccacacgtcacccgctgagcagcagcagcagcagcagcagcagcag 1206
|||||
Db 17088 CACAGCAAGCGCCACCGCTGACCGCTGACCGACCGGACCGACCGACGATCACCCGAG 17029
QY 1207 ctgacctggcctggcgccccacacacgactgcgcgaaaggctggaccacccacac 1266
|||||
Db 17028 CTGACCTGGCCTGGCGCGCCGACACCGGCTGCGCGAAAGGCTGGACACCCCAAC 16969
QY 1267 aacacccagccacacccgaatggtaccacacccacccctgacccatggccaccccg 1326
|||||
Db 16968 AACACCCAGCGGCACACCGAATGGTTACCAACCAACCCGACCTCGACCGCCCAACCGTGG 16909
QY 1327 accaacaccttcacaccccc 1347
|||||
Db 16908 ACCTGTGAGTACACTACAC 16888

RESULT 8
MTCY251/c 38380 bp DNA BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 5/162.
DEFINITION 274410 AL123456
ACCESSION 274410.1 GI:3261600
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium: Mycobacterium tuberculosis complex.
1 (bases 1 to 38380)
Cole, S.F., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,

Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Szares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (5685), 537-544 (1998)
98295987
2 (bases 1 to 38380)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1405752.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
location/Qualifiers
1..38380
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
..35889
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
..956
/clone="Y251"
6..956
/gene="hycB"
6..956
/gene="hycB"
/note="KV0084, (MTCY251.02), len: 316, hycD, similar to
HYD_ECOLI Phd40 formate hydrogenlyase subunit 4 (307 aa)
FASTA scores, opt: 570, z score: 591.6, E(): 2.1e-26,
(4.8% identity in 405 aa overlap) and NUCL_ECOLI
P460 nadh dehydrogenase i chain h (325 aa) FASTA scores,
opt: 207, z score: 220.7, E(): 9.5e-06, (26.5% identity in
260 aa overlap)"
/codon_start=1
/transl_table=1
/product="hycB"
/protein_id="CAA98920.1"
/db_xref="GI:1405754"
/db_xref="SPTREMBL:O10881"
/translation="MSYLACAAOIGVMVGVAPLVIGMTROVRARWLCRAGAGLLQPWR
LLKQLGKQKITPAGTIVFAAPVIVAGTTLIAAPLAVTSGPDDPSADLFAVVG
LLELGYVALTAGIDTGSFGMGASRGTIIALVEPTILLAVFALSIPAGSALGAL
VASTIDHPGHVSLAGVLAPVALVIVVAVTGRPDVNPATHLTMTVMHEAMVLEYAG
PRALIVEMAGMBRLTVLLALLFLPWGIAAGAPTALDVLTCGWAAAKVA1LAVLL
ATFEVFLAKLRLFRVPELLAGSFLALLAVTAANFTTGA"
967..1629
/gene="hycP"
967..1629
/gene="hycP"
/note="RV0085, (MTCY251.03), len: 220, hycP, similar to
spiP7524|HYFE_ECOLI HYDROGENASE-4 COMPONENT E (216
aa), FASTA scores: opt: 204 z-score: 272.1 E(): 1.2e-07;
25.5% identity in 216 aa overlap"

```
/codon_start=1
/transl_table=11
/product="hycp"
/protein_id="CAA98921.1"
/db_xref="GI:1405755"
/db_xref="SWISS-PROT:Q10882"
/translation="MSNAFSLVDFPAGGLVLASVLIVRRDLRAIVRLLANQGAAL
AAILPGRINDRALJAVGIALLALRALVLPWLLARAVGAARAAQREATPLVNTASS
LLITAGITLFAITQPVNLEPGVTINPAFAVVLIALFVMTLRLHAVSOAAGFL
MLDNGIATFLTAGVPLLIVELGASLDLVFAFIVIGLVLTGRLLRIFGDADLDKREL
RD"
gene 1629 .3095
CDS
/gene="hycQ"
1629..3095
/gene="hycQ"
/note="Rv0086, (MTCY251.04), oxidoreductase, len: 488,
hycQ, most similar to sp|P77437|HYFF-ECOLI HYDROGENASE-4
COMPONENT F(EC 1.- (526 aa), FASTA scores: opt: 948
z-score: 1117.4 E(): 0; 35.9% identity in 493 aa overlap.
Also similar to E. coli d9087711 & NUOL-ECOLI P33607 nadh
dehydrogenasel chain 1 (613 aa) FASTA scores, opt: 360,
z-score: 354.9 E(): 3.2e-13, (27.9% identity in 488 aa
overlap), and to NUON-ECOLI P33608, nadh dehydrogenase 1
chain n (425 aa), FASTAscores, opt: 375, z-score: 371.4,
E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/db_xref="SPTREMBL:Q10883"
/translation="MTGLLAILAPLAAIASLITCWRRRTTATLALSATTVLACAV
AMGFVMSGAQGLGLLRADALTVMVLVIGIVGTALTAASIGYIDTELHGHIDGR
SARLYGVLPAPICAMVLAVANNIGVIVAEITVITAFVGHRRPTALEANTKY
VVICSVGIAPFTGLVLYFAARDSCAAAGALNLDIAEHAGLDPGVARAGGLLL
IGVCAGAGLPFHTLADASHQAPAPVSLMSGVLLAVAFILRLRPLDVAVGPAV
LRNGLLVGLTLLVALLMLTITVDKVRMLAYSSMHMGLIAIAAAGTTAIAALLL
HVLHGIKTVLFLAGLOQAAHDSIAIDITGVMRRLSIGVSAFVGLIVLLGLPPF
AMFASLARSLSANLAWLWGLAALLLIAGFTALARNSGRMLLTGPAAGPATVTP
ATAAALMVGIVVSAALGITAGPLADLLGLAASNVGLP"
gene 3092..4570
CDS
/gene="hycE"
3092..4570
/gene="hycE"
/note="Rv0087, (MTCY251.05), oxidoreductase, len:
492,hycE, similar to HYCE-ECOLI P16431 formate
hydrogenlyase subunit 5 (569 aa) FASTA scores, opt: 680,
z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa
overlap) and to NUOD-ECOLI P33600 nadh dehydrogenase 1
chain d, (407 aa) FASTA scores,opt: 245, z-score: 293.1,
E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/db_xref="SPTREMBL:Q10884"
/translation="MWSASLHRHVSERGLIATAEQLWADSFRLVAHDDGDSLRV
VYFLAGYDPRVELEYVVPADNPEIRSLAYLSFPAGFRFEMADLYGIRVPVGHKPR
RLVRHAGMDHPMRTDAGPFEFTDGFAPFLAVEGCVVEIIPVPGVHAGLIEPHF
RFSVAGETIVRLKALWFWHVGIEKLFHCRPATAVDLAERIISGDTSAHAHLISLAI
EDALGIELPHEVHRLALIVELRYLNNHADLGALANDGVYSLANAHQRIKENLLRI
NAAVTGHLRCAIRAGGVALRALPDDTDELAALVDLAELATLTLANSVYVDRFATGA
VLPDDASALGCGTVARASGLRSDARVEHPTIVLPITEIGAPDGDVLARTVRRDEF
AASALAQHI VESHPTGPIEYATLHPVGAPSSGIGIVEGWRGTIVHVRVEIDVDGRIIR
ARKVDPDSFNFPALPVAMADTIVDPFPLANKSFNQSYAGNDL"
gene 4605..5279
CDS
/gene="Rv0088"
4605..5279
/gene="Rv0088"
/note="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
```

```
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/db_xref="SWISS-PROT:Q10885"
/translation="MSVYKHPFSLRVLRTSTRVTVKVRSGSLSWRRVRTGDLGLAVMG
GREYRAVAPGTPGIPQKGMMTVTIVDAGPGRVSRSEVAAAEALFAIVADPRRHR
ELDGSSTVRGNIKPAKLVGSKFTKMKLFGLPYRITSRTVTKLPHNELVESHPLGH
RWKEPESLSPTLTRVTEFDYHAAGAIAKNGLKVFYEMTGFAKSNAAAI EATLAKLSQD
YARGRA"
gene 5436..6029
CDS
/gene="Rv0089"
5436..6029
/gene="Rv0089"
/note="Rv0089, (MTCY251.07), len: 197. unknown, some
similarity to sp|P12999|BIOT-ECOLI BIOTIN SYNTHESIS
PROTEIN BLOC. (251 aa), FASTA scores: opt: 202 z-score:
262.0 E(): 4.5e-07; 39.0% identity in 118 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0089"
/protein_id="CAA98925.1"
/db_xref="GI:1405759"
/db_xref="SWISS-PROT:Q10886"
/translation="MDQPNANIHYDALLDAMVPLGTCQVLDVGGDGLLAARLARRI
PYVTAVDDAPVLRRAQTRFANAPIRLHADIMTAELPNAGFDVAVSNAALCHIEDTR
TALSRLUGGLVTPGGLTAVVTFVTPSLRNLWHLTSWVACGMANRVKKGWHSAPIKWP
PPOTLHELRSVHRALLPGACIRRLLYGRVLVTRAPV"
RBS 6147..6152
/gene="Rv0090"
6158..6928
CDS
/gene="Rv0090"
6158..6928
/gene="Rv0090"
/note="Rv0090, (MTCY251.08), len: 256. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0090"
Query Match 93.4%; Score 1301.4; DB 1; Length 38380;
Best Local Similarity 96.6%; Pred. No. 1.5e-168;
Matches 1340; Conservative 0; Mismatches 1; Gaps 1;
QY 7 cgggtagccgcgaacgattgtcgaggcttttgatgcgctgatatccgagctggaccct 66
Db 12772 CGAATAGCCGGAGGAGCTTTCGAGGCTTCGACGCGCTGGATGCCACCTTGGACCGCT 12713
QY 67 tggacgaagtgatgtttggaatgtttgacatccccaagacgctgctgtcttggaacgtc 126
Db 12712 TGCAGCAGCTGTCTTTTTCAGCTGTTCAGCAACCCGCTGCGGCTCTCTGGAACCTC 12653
QY 127 tgaatgtttgttgcgcgctgttcggttgggttcacgtttgtatcaacaaactllqacg 186
Db 12652 TGAATGTTCGTCGCCCGCTTACCGGCTTCCACACACACAGTTGATCAACAACTCGACA 12593
QY 187 cccaagcagcaggaagaactggcgacgctgtgtgcgcgctggccaccggttac 246
Db 12592 CCCAAGCAGCAGGAAGAACTGGCGCGCAGCTGTGTGCGCTGCCAACCGGCTGC 12533
QY 247 gcatcacaagcccgacgcccgcgcacatgcgcgacgcccgcgcacatctcggaacctctc 306
Db 12532 GCATCACAAGCCCGAAGCCGCCCGCACGCGCCGCAAGCC-AAGCCCTTAGGGCAACGTC 12474
QY 307 gagcactcaccggttgaaacgcttagccacacagttgaccgcccaccgcccaccgccc 366
Db 12473 GGGCGCTCACTGGCGCAACCGTTGGCCCCCGCAGTTGACCGCCACCGCCGCCAACGCC 12414
QY 367 agggcctgacgagcagcagctcaagtgtattcgcccccttttgcgccacctgccc 426
Db 12413 AGGGCCTGATCGCGAGCGCAGCTCAAGTGTATTTCGCCCTTTTTCGCCCACTGCC 12354
QY 427 gcccggttgatgttccaccgcccagccgacgacgacgacgacgacgacgacgacgac 486
Db 12353 GCCCGGTGGATGTGTCACCAGCCGCCGGAAGCCGACCTGGCGCGCAAGCCGCT 12294
```

[illegible]

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNQNRINRWELITCGLGHHVTPADDAALAARLARASTGLGE
VWRKLCGGDFALGPGGPGGADAPLIMRGKALRQRIIRALCVERLPRALVLAALAA
AVMEFRGAAGIQAOTLDRDLPLVLRAGAFVQDMQTVIHALEKAKALKPSTLALITGHA
AYAVLQAQVEGVGLWLLKRWGEYFAVYATIFLPLEVHDLAKGKITTRVVTFVTSINVAAY
VYLLISKRLFVGRGRKAYDVERRGQLDLERAAFLT"
3791.4162
/gene="MT0099.1"
3791.4162
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MRGLRVVSKVTSFALFTPPARRAKAAHVNLDOVAOCERTDEG
PTLCOHQPGSARALPTAAWSRQSRVPATHCRPPCAPGAASALTCAEAWSVVE
VRPAPRAETSQASRAARESR"
4206.4973
/gene="MT0100"
4206.4973
/gene="MT0100"
/note="similar to GB:ALC009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine
nucleosidase/S-adenosylhomocysteine nucleosidase,
putative"
/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTVGVICAIPOELAYLRGLVDKARQOQVLAIFDSGGQLDHRH
VLLAAGMKVNTGILATLADRGGCRITVFTGVAGGLDPCLGIDGIVIAADVVRQHDIF
GLLDELRQVGHGPIFIEPTERLGVDPDVIDRKHLRDLGTFLAPLSTAAGGGGR
QPRVYGTILTGDDYLLHCERTNRLLHLELGMVAMEGGAVAQICASFIDIPWLVIRAL
SLDLAGQSGVDNFRFVGEVAASSARVLRLLLPVLTAC"
5105.7390
/gene="MT0101"
5105.7390
/gene="MT0101"
/note="similar to SP:P46840 PID:559912; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTAVTCEHHASVQIQLRISGMSCSACARHVESTLNLKLPGRVRA
AVNFRVATIDTSEAVDPAALCOAVRRAGYQADLCTDDGRSADPDADHARQLLIRL
AIAAVLFPVADLWSMGEVVPATFTCWQGVLSALAPVVTWAAMPFRHVRMNRARHIL
AASMETLISVGIATATWLSLITVGNHSPTERSGIWOALLGSDAIFYEVAAGVTVFVL
VGRYFEARKSQGASLRALAAALSAKEAVLLIPDGSEKVIIPADELKKQQRFPVRPGQIF
VAADGLVDGAAVDMSAMTGEAKPTVRPQGVQVIGTVDLDRGLIVEAAAGVADTQFIR
AGWVLRPQAQQAQADARADRISSVFPAVLVIAALTAAGWLIAGGQPDRAVYSAAL
AVLITACALGLATPTAMMVASGRGAQLGIFLKGYSLEATRAVDTVFDTGTLTAT
GRLOYSAVTAAPGWEADQVLAALVBAASHSVALAIAAATTRRDVATDFRAIPGRG
VSGTSGRAVVRGEPKSHLSSSCSHNMRARHAEISLGETAVFEVDGEGPCGVLAADV
AVKDSARDAALADRLRLMTLLTGDPNPESAANAATVRGIDEVIADILPRGKVDVIEQ
LDRGHVYVAGWDGINDGFPALARADLGAIGRDTLAAAGLNLPLVAGAAMAFSSFPVNSLRLR
LARATMTVKYKLVNWYGVGATIPVAAGLNLPLVAGAAMAFSSFPVNSLRLR
GRYPUGCTGSGGQPMTPASSA"
complement(7337.8185)
/gene="MT0102"
complement(7337.8185)
/gene="MT0102"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

<pre>/protein_id="AAK44325_1" /db_xref="GI:13879149" /translation="MLAQATVAGSPNHHASTVLQCGRGVPAAWSEPAARHRCATYI DMDCEVARREALSRADGERAPVPSRVDEHLGECSSACRAWFQTVAHQGDURKLAS RDPWPVGRLGTTRRAPRRGHSTPWRWALLCVGIAIQIALGTQGGLVDGLTHQHPT GAGTTLHNSTWSIALGVIMVGAALWPSAAAGLAGLTAFAVLLTYIVDALSCAV SVTRLTLPLPVIGAVLAIMVRSASGRPRPDVAAEPIDIVLPDNASKRKRGHLMP TNGSAA"</pre>					
gene	complement(8232..9095)				
CDS	/gene="MT0103" complement(8232..9095) /gene="MT0103" /note="identified by Glimmer2; putative" /odon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK44326_1" /db_xref="GI:13879150" /translation="MDVIHFDDGLDTUTERAKRKGITLSNOOYDCMSRLSGYLTPQARA TKLFVAILKAAPCAINPDHPITVDIPDAEAIDRTDRSQARNHNDGLILAGILRALIAS GKIQLHGHLNPSSIVVTITITLDLTGAGKGCTGGTLLPMADVIRMTSHAHUYSPASGR YQAIFDHGTPLALYHTFKLASPAQRIMLFANDRGCTKPCSDAPAYISOAHHVTAWTS TGHDITEFTLACCPNRLAKEGWTHKNTHGHTEWLPPPHLDHGQPRTNTFHIUPERE LIHQDDDDKPD" complement(8956..9623) /gene="MT0104" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; identified by Glimmer2; putative; conserved hypothetical protein, authentic frameshift"				
BASE COUNT	1553 a 3103 c 3360 g .1748 t				
ORIGIN					
Query Match	93.2% Score 1298.2; DB 1; Length 9764:				
Best Local Similarity	96.5%; Pred. No. 7,7e-168;				
Matches 1336;	Conservative 0; Mismatches 48; Indels 1; Gaps 1;				
Qy 7	cgggtagccgcgaacgattgcaggcttttgatcgctggatgccagctggaccgtc 66				
Dz 9616	CGAATAGCCGCCGAGAGCTTTCGAGGTCCTTCGACGCCTGGATGCCACCTGGACCGCT 9557				
Qy 67	tggacgaggtgtcttllgagtggttgaacccccagagaagctcggtctcttggaaactc 126				
Dz 9556	TGGACGAGGTCCTCTTTGAGTGTCGTGACGACCCCGGAACGGCTGCGGTCTCTGGAAACTC 9497				
Qy 127	tgggaatgttgttgcgcgcgcgtacctgagtcgggltgggtctgatcaacaacttggacq 186				
Dz 9436	TGGAATCTTTTCGAGGAGAGGCTAACGGCTTACAGGCTCATGATTGATCAACCAACTCGGA 9437				
Qy 187	cccgaagcagcaggaagaaataatctqgcgcgcacgtctgctgycgcttgccaaactggttac 246				
Dz 9436	CCCAGGCCAGCGAGAAGAACTTGGCGCGCACGCTGCTGCTGCGCGCTGGCCAACCGGCTGC 9377				
Qy 247	gcataccaagcccgacgcccgcgcgcacatcgccgcgcgcgcgcgatctcggaactcttc 306				
Dz 9376	GCATCACCAGCCCGCAACCGCGCCGACGCGCGCCGAAGCC -AAGCCTTAGGCCAACACTC 9318				
Qy 307	gagcaactcacgcgtlgaaacgcgtlagccccacagttagccgcacccgccacccgcacgcgc 366				
Dz 9317	GGGCGCTCACTGGCAACCGTTTGGCCCCCGCAGTTGACGCCACCGCCACCGCCCAACGCC 9258				
Qy 367	agggcctaactcgcgagcgacgcacgtcaagtattcgcccttttgcgccactgcgc 426				
Dz 9257	AGGGCCTGATCGGCGAGCGGCACATCAAAGTGATTTCGCGCCCTTTTTCGCCCACTGCC 9198				
Qy 427	gcgcgcgttgatgtgtccaccgc 486				
Dz 9197	GCCGCGGTGGATGTCTCACCGCT 9138				
Qy 487	caatatgtccgcagcagctggccgcgtacgcgcagcgggtcatgactggctacacccc 546				
Dz 9137	CAATAATCGTCCCGAGAGCTGGCCCGCTACGCCCGACGGGTCATGGACTTGGCTACACCCC 9078				

Qy	547	gacgcgacctcagcacacgaacgaacgccccgcaaaacggcgatcacccctgagcaaacag	606
Db	9077	GACGGGACCTTACCGACACCGACGCGCCCGCAACGCGCATCACCTTGAGCAACGAG	9018
Qy	607	caatacagcgcgatgtcacggctgaagtgtgctacctgaccccccaagcgcgagccacttt	666
Db	9017	CAATACGACGCGATGTACGGCTAAGTGGCTACCTGACCCCCCAAGCGCGGCCACCTTT	8958
Qy	667	gaagcgcgtgtctagcaaaactggcgcccccgcgcgacacccaccccgagcacaccccg	726
Db	8957	GAAGCGGTGTAGCAAACTGGCGCCCCCGCGCGACCAACCCCGAGCACACACCCCG	8898
Qy	727	gtcatcgacacacccccgatgcgcgccatcgacgcgacaccccgacgaagccaa	786
Db	8897	GTTCATCGACACACCCCGATGCGCGCCCATCGACCGGCACCCCGACGCCAAGCCCAA	8838
Qy	787	cgaacacacgacggctgtgccccggcgcgctgctgacccctgacccctgacccctgacaaac	846
Db	8837	CGCAACACGACGGGCTGTGGCGGGCTGCGCGCGTGTATCGCTCCGGCAACTGGCG	8778
Qy	847	caacacacggtcttcccgctcgatcggtggtcaccacacccctgacccctgacccctgacaaac	906
Db	8777	CAACACACGCTTCCCGTCTCGATCGTGGTCAACACACCTTGACCGACCTGCAAAAC	8718
Qy	907	ggcgccggcgaaggcttcaacggcgcgccacccctgctacccatggcgcatgltgatccgc	966
Db	8717	GGCGCGGCAAGGGCTTACCGCGCGCGGACACCTGTACCCATGGCGCATGTGATCCGC	8658
Qy	967	atgacacgacacccacactactcccgcaagcggaaggtacccccaggcgatcttc	1026
Db	8657	ATGACGACCGACCCCACTACTCTCCCGCAAGCGGGAGGTACCCCGCAGCGCATCTTC	8598
Qy	1027	gaccacggcacacccctggcgctgtatcacaccaaaccgcttagcctccccggccccagcg	1086
Db	8597	GACCACGGCACACCCCTGGCGCTGTATCACACCAACGCTAGCCTCCCGCGCCAGCG	8538
Qy	1087	atcat..tggtccgaacagccgctgacacaaaccccgctgtgacgcacccgacctac	1146
Db	8537	ATCATGCTGTTCGCGAACGACCGCGGTGACCAAAACCCGGCTGTGACGACACCGCTAC	8478
Qy	1147	cacagcaagccccaccacgtcacgcctggacacgacacgacgacgacacacccgag	1206
Db	8477	CACAGCAAGCCACACAGGTACCGCTGGACACGACCGGACGACACCATCACCGAG	8418
Qy	1207	ctgacctggcctggcccccgacaaccgactcgcgaaaaaaggctggacacacacac	1266
Db	8417	CTGACCTGGCCTGGCGGCCCGACACCGACTCGCGGAAAAAGGTGACCAACCAAAA	8358
Qy	1267	aacaccacggcacacccgaatggtaccacacccacccctcgaccacgacacacccgac	1326
Db	8357	AACACCCAGGCGCACACCGAATGGCTACCAACCCACCTCGACCACCGGCAACCCCG	8298
Qy	1327	accaacaccttcaacacccccgaagattcttccacaaacgaagacgacgacacaccc	1386
Db	8297	ACCAACACCTTCCACACCCCGAAGATTCTTCCACAAACCAACAGACGACGACGACAA	8238
Qy	1387	gattgac 1393	
Db	8237	GATTGAC 8231	

RESULT	10
MBDR3S1	
LOCUS	MBDR3S1 9281 bp DNA BCT 27-APR-1996
DEFINITION	Mycobacterium bovis deletion region 3, 5' end.
ACCESSION	U35017
VERSION	U35017.1 GI:1049238
KEYWORDS	1 of 2
SEGMENT	
SOURCE	Mycobacterium bovis.
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium tuberculosis complex.			
REFERENCE	1 (bases 1 to 9281)		
AUTHORS	Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.		
TITLE	Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis		
JOURNAL	J. Bacteriol. 178 (5), 1274-1282 (1996)		
MEDLINE	96200095		
REFERENCE	2 (bases 1 to 9281)		
AUTHORS	Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology, Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA		
FEATURES	location/Qualifiers		
source	1..9281		
	/organism="Mycobacterium bovis"		
	/db_xref="taxon:1765"		
	misc_difference replace(1400..9281,"")		
	/note="absent in related avirulent strain M. bovis BCG"		
BASE COUNT	1706 a 3191 c 2929 g 1453 t 2 Others		
ORIGIN			
Query Match 92.0%; Score 1281.8; DB 1; Length 9281;			
Best Local Similarity 97.2%; Pred. No. 1.3e-165;			
Matches 1304; Conservative 0; Mismatches 37; Indels 0; Gaps 0;			
QY	7	cgggtagcgcgaacgagattgtaggttttggatgc--trgatccgagctggaccgct	66
Db	84	CGAATAGCGGGAGGAGTGTTCGAGGTTTGTATGCGCTGGATCCGAGCTGGACCGCT	143
QY	67	tggacgaggtgtcttttgaaggttgaccaccccaagaacgctcggtctcttggaaacgc	126
Db	144	TGGACGAGGTGTCTTTTGGGTGTTCACACCCCGGACGGCTCGGGTCTCTGGAACGTC	203
QY	127	tggaaatgcttggcgccggctacacgqcggtgggttcacgcttgatcaacaaacttgacg	186
Db	204	TGGAATGCTTGTGCGCGGCTACCGGGTGGGGCACACGTTGATCAACCAACTCGACA	263
QY	187	cccaagccagttacaaactggtggtgacgctgctgtctgctgctgctgctgctgctgct	246
Db	264	CCCAAGCCAGGAGGAAGAACTTGGGCGGACGCTGTGTGCGGCTGGCCAAACGGTTAC	323
QY	247	gcatcaccacgaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac	306
Db	324	GCATCACAAGCCGACGCCGCCCTTACGATCGCCGACGCGCGGATCTCGGATCTCTC	383
QY	307	aadgaactcaatggtgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat	366
Db	384	GAGTACTTACGATGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT	443
QY	367	agggcctgacggtgagcgcaatgtagtgcgccccttttgcgacacgctgac	426
Db	444	AGGGCTGATCGCGGAGCGCACATCAAGTGTTCGGCCCTTTTTCGCCACCTTGCCC	503
QY	427	gcccgggaggatglttccaccccgacgacgacgacgacgacgacgacgacgacgacgacg	486
Db	504	GCCCGGTGGATGTGTCCACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGCT	563
QY	487	caatatgctcccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac	546
Db	564	CAATATGCTCCCGACGAGCTGGCCCGGTACGCCGCGGGGTATGGACTGGCTACACCC	623
QY	547	gacggcacttcccgacacacgacgacgacgacgacgacgacgacgacgacgacgacgacg	606
Db	624	GACGGCACTTACCGACACCGGAACGCGCGCGGCAACCGCGCATCACCTGAGCAACG	683
QY	607	caatacagcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac	666
Db	684	CAATACGACGCGATGTACGGCTAAGTGGTACCTGACCCCGGCGGCGGCGGCGGCTTT	743
QY	667	gaagccgttctagcgaactggcgcccccgcgcgacaccccgacgacgacgacgacgacg	726


```
gene
complement(3286. .3654)
/gene="Rv1137c"
CDS
complement(3286. .3654)
/gene="Rv1137c"
/notes="Rv1137c, (MTC165.04c), len: 122. unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1137c"
/protein_id="CAB09028.1"
/db_xref="GI:2117175"
/translation="MLSARHIRHSPGDKARCAHLSATLRPGIGISPTNVGNATVL
ADGTPAKPIQGAETMORARHTGSCFSANARGPALSSGNPSRAGCGVSSITTSPTQ
AIRLLACTSDALVTYTR"
complement(3659. .3662)
/notes="possible RBS, AGGA, for Rv1137c"
complement(3671. .4687)
/gene="Rv1138c"
complement(3671. .4687)
/gene="Rv1138c"
/notes="Rv1138c", (MTC165.05c), len: 338. Similar to many
M.tuberculosis hypothetical proteins e.g. TR:P72008
MTCY04C12.35 (412 aa), fasta scores; opt: 89 z-score:
292.1 E(): 4.5e-09, 24.6% identity in 358 aa overlap.
TBparse score is 0.929"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1138c"
/protein_id="CAB09026.1"
/db_xref="GI:2117176"
/db_xref="SPTREMBL:O06538"
/translation="MPSYDPLLVGGPGGLATALHARAGLSIVAEPRNPIDKA
CGEGLMPGGLAEITSLGVDPVGFPHGIATVGEHRRVQARFTGPGRGVRRITLHAAL
AARAKEDTWRISRVATIQDADHGVTAAGVRAKWLVAADGLHSVRRVAVGIKATAGT
PRYGVHMYRLDSVDFEVHWSRMGEAYVTPEPDLVGVAILSRQRPPLAFWPSLA
HHLQDASRGHARGCGPLRVQVSRVAGRVLLVGDAAGYEDALTGEGISLAVKQAAAV
SAIVDDTPASVAAWHRITRDYRLVTRGLVLAATPRAARRAIVPLCALLTATFRYGN
ILAY"
complement(4684. .5184)
/gene="Rv1139c"
complement(4684. .5184)
/gene="Rv1139c"
/notes="Rv1139c, (MTC165.06c), len: 166. Highly similar
to hypothetical B. subtilis protein, YHPO. FASTA scores:
YBQ_BACSU HYPOHETICAL 19.3 KD PROTEIN IN BCSEA-DE (168
aa) opt:446; E(): 2.2e-26; 38.4% identity in 164 aa
overlap. TBparse score is 0.911"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1139c"
/protein_id="CAB09027.1"
/db_xref="GI:2117177"
/db_xref="SPTREMBL:O06539"
/translation="MYLLILAVFERLAELVVAQRNARMSFAQGGKEFGRRHYVVMV
ILHTALLGCEVPWALHPRFPILWGLMPLVAVSOGRLWMCVKSLGRKRNTRIVL
PHATLVRRGPRVMRHPNVAVVAEGFALPLVHTAWLTALVETLANATLLTVLRIVEN
SVLGYT"
complement(5192. .5197)
/notes="possible RBS, AGGTGG, for Rv1139c"
5546. .6394
/gene="Rv1140"
5546. .6394
/gene="Rv1140"
/notes="Rv1140", (MTC165.07), len: 282. Weak similarity in
C-terminus to hypothetical E. coli proteins YPRA and YPRB,
possibly membrane-bound. FASTA scores: YPRA_ECOLI
HYPOHETICAL 24.3 KD PROTEIN (URF 1). - (217 aa) opt:
166; E(): 0.00062; 31.0% identity in 158 aa overlap.
TBparse score is 0.927"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1140"
```

```
/protein_id="CAB09028.1"
/db_xref="GI:2117178"
/db_xref="SPTREMBL:O06540"

Query Match      50.2%  Score 699.6  DB 1:  Length 34331.
Best Local Similarity 70.4%  Pred. No. 7e-87:
Matches 975:  Conservative 0:  Mismatches 384:  Indels 25:  Gaps 2:

Qy 1  tgggtttcgggttagcgcgcgaacggatlytgcgaggtcttltgagctggatgcgctggatgcgagctgg 60
Db 15862  TGAGTGTGGACACCGCGCAGCAGATCTCCCGCGGTGTGGATGGCTTACCACGCCCTGTGT 15803

Qy 61  acgcgttggacgagtgctcttlltgaagtgltlgacaccccccaagaacggctgcggtcctctg 120
Db 15802  CGCGGGTCTCTCGATCTCAAGTGCATGCGTGTGACCAACCCCGGAATTGCTGGCCCTGTTTGC 15743

Qy 121  aaqctctggaaatgcttltggtgcgcgcgttaccgcgcgltgggtcacgcgttgaataaaccaac 180
Db 15742  AGCGACTCGAGCTTGAAGGGGCGCGCAGGTCGCGGTCGACGACGCCCTTGATCAACCAAC 15683

Qy 181  ttgacgcctcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
Db 15682  TCGTGTGGGCAAGCTTGCAGCAGCAGAGCTTCGCGGGAGCTCGCGCAGCGCTTGCCCAACC 15623

Qy 241  ggttacgcacatcaacaaqccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 15622  GGCTACACATCATCTCCCGGTGAGGCCAGCGCGCATCGCCGAAGCGGAAGACCTCGGTG 15563

Qy 301  ctgctgagcacttaccgcgtgaacgcgttagcccaacagttgacgcgcacgcgcgcgcgcgcgc 360
Db 15562  AGCGCGCGCCCTGACCGGTGAACCGTGTGCGAGCGCAGTTGACCGCGACCGCGCGCCGCTC 15503

Qy 361  aacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 15502  AACGTGAGCGCAAGATCGCGCGAGAACACATTAAGAGAGATCCAGG-CCTTCTTCAAGCAG 15444

Qy 421  ctgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
Db 15443  TTGTTCGCCCGCGGTGATCTGGGTATCCGCGAGCCCGCGAGGCCAGCTGGCGCGAACTG 15384

Qy 481  gccgcctcaatctatctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 15383  GCCACCACTGAGATCTCCCGATACCTGCTATGCTGCGCACGACGATGATGGATGGCTG 15324

Qy 541  caccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 15323  CAGCTGAGATGAAATTTCTGTAATAGATGTCGCTGCAAGCGCGCATACATCATGCT 15264

Qy 601  aaccgcgaatcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660
Db 15263  AAGCAGCAATTCACCGGATGTACGATACGATGATGATGATGATGATGATGATGATGATGATG 15204

Qy 661  accttgaagccglgttagcgaactggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 15203  ACCATCAGCGCGTGTGTGGCCAAACTGGCGCGCACCGGGGCGGTGCAAGCCCGATGACAG 15144

Qy 721  accccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 15143  ACCCCGCTCGTGGATGACACACCGGATGCGGACGCGGTGCGCGCGACACACCGCGACCG 15084

Qy 781  gcccaacgcgaacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 15083  GCACAACCAATCATGATGCGTGTGCGCGCGCTGCGCGCGCGCTGTTGGCTTCGGGTGAG 15024

Qy 841  ctgggcgaacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 15023  CTGGGCAACACAAAGGGGTGCGGTAAACGATCGTGTGTGAGCAGCAGCTAAGAGAGCTG 14964

Qy 901  caaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 14963  GAAGCCGCCACCGCAAGGGGTAAACGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14904

Qy 961  atccgcgaatgaccagccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
```


Db 9856 GCACACGCAATCATGATCGTTTTTGGCCGCGTGCAGGCGCTGTTCGCCTCCGGTCAG 9797
QY 841 ctgggccaacacacggcttcctccgtctcgcgtcgtggtcaccacaccctcagccagctg 900
Db 9796 CTGGGCCAACACACAGGGGCTGCCGTAAACGATCGTGTGGTGACACACACCTAAAGGAGCTG 9737
QY 901 caaacggcgccgggcaagggtctcaccggcgccggcaccctgctaccacatgcccgtg 960
Db 9736 GAAGCGCCACCGCAAGGGGTAAACGGTGTGGTTCGCCGGTGCCTGTCGACCTT 9677
QY 961 atccgcatgaccagccacgcccacactactcccccgcaagcgaggtacccccagcgc 1020
Db 9676 ATCCG:ATGGCGAGCCACGCCAACCACTA-----TCTGCCA 9641
QY 1021 attctgcacacggcacaccctcgcgtgtatcacaccaaaccctagctctccccggcc 1080
Db 9640 TTGTTTACGGCGCTAAGCCCTTGGCGTTGTATCACCAACCGGTAGCTTCCCGCGCG 9581
QY 1081 cagcggtatcgtgttgcgaacgacggcggtgcaccaaaccggtcgtgtagcgcacgc 1140
Db 9580 CAGCGAATCATGTTGTACGCGAAGGATCGTGGCTGCTCCAGGCGGCTTCCGACGCCCG 9521
QY 1141 gcttaccacacgaagccacacgctcaccgctgacgcgctggccgacccgacgcacgacatc 1200
Db 9520 GCCTACCACTAGTGGTCCACCGCTAACCGCTGGGACAAACCAACCGCTACCGACATC 9461
QY 1201 accgagctgacctggcctgcggcccgacacgacgctgcgacccacccctcagccagccaa 1260
Db 9460 AACGACCTACGCTGGCTGCGGCCCGACCAATCGCTTGTGCAAAAGGCTCGAAGACC 9401
QY 1261 cacaacacacacgacgacacgaatggtcaccacacccacccctcagccagccgcca 1320
Db 9400 CGCAAAACGCCCGACCGACCGAATGCTTACCAACCAACCGCTCGACCAACCGCCAA 9341
QY 1321 cccgcgacacacaccttcaccaccccggaacgacgattctccacacaaacgacgacgac 1380
Db 9340 CCGCAATAATCATGATACCAACCGCCCGCAATCTGTGCGCAACGACGACGCGAA 9281
QY 1381 aaac 1384
Db 9280 CCAC 9277

RESULT 14
MTCY9F9
LOCUS MTCY9F9 33100 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 87/162.
ACCESSION Z84498 AL123456
VERSION Z84498.1 GI:3261701
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekla, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REFERENCE 2 (bases 1 to 33100)
AUTHORS Parkhill, J.
TITLE Direct Submission

JOURNAL

COMMENT

FEATURES

Source

source

gene

CDS

misc_feature

gene

CDS

Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1806198.

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers

1..33100
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/cl..32559
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y9F9"
265..1380
/gene="Rv1922"
265..1380
/gene="Rv1922"
/note="Rv1922, (MTCY09F9_42c), len: 371 aa. Similar to penicillin binding proteins and eg DAC_STRSQ P15555 d-alanyl-d-alanine carboxypeptidase (406 aa), fasta scores, opt: 382.E(1): 3.1e-17, (28.0% identity in 379 aa overlap); containsPS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1922"
/protein_id="CA050504.1"
/db_xref="GI:1806240"
/translation "MISTVTFASIRRMRLGLIAATLLGCGCTGQHTTTAASTVTPHIK ASSQDVLDIAINAHPECTSAAGVGRKVIWGVRG IADLASGAK IFTDTVFDIASYSK QETATAILLVFAAGKLTIDPIISUYVELPDMAQTVTYEQLMHQTSGLTIDYVALLAAR GYVUSDRITKAEAKAALAAPELOFKPGRFDYNSNLYLLGEIVHRASGQPLPEFLS AEIPLQGLANVYDVGKVPKNKAVSTYKGTGGRNRRSEYRVGNPAWEQIGDGGIQTTPSQ IARWADNRTGSGVGIKLLAQLAGAVETEPGGDRYGAGIVSRADTLDHAGAWAGF IAFHISDDRRTSVAISCNVDKPDVPAMADALGRLWM"
307..339
/gene="Rv1922"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
1371..2711
/gene="lipD"
1371..2711
/gene="lipD"
/note="Rv1923, (MTCY09F9_41c), len: 446 aa. Similar to esterases, beta-lactanases, eg GL51214 esterase, (389 aa), fasta scores, opt: 569, E(1): 5.4e-29, (33.7% identity in 401 aa overlap) also MTCY277.19 (56.6% identity in 410 aa overlap)"
/codon_start=1
/transl_table=11
/product="lipD"
/protein_id="CA06503.1"
/db_xref="GI:1806239"

Db 13128 TTGTCGCGCGGTGGATCTGGGTATCCGCGAGGCGCGGAGGCCAGCTGGCGGAAC TG 13187
QY 481 gccgctcaatagtctccgacagagactggcccgctacgcccagcggtcatggaactggcta 540
Db 13188 GCCACACACTCGCGCTCCCGATCACCTGCTATGCGCTGCGCCACGACGCTGATGAGCTGGCTG 13247
QY 541 caccgccagcgacccctcaccgacacccgaaacgcccgcgaacacgcggtcatcaccctgagc 600
Db 13248 CACCCCGACGCAACTTTTCGACACGAGGAGGTGCCCGCAAGCGGCGGATCACGATGGGT 13307
QY 601 aaccagcaatagcagcgatgtcagcgctaaagtggctacgtgaccccccaagcgcgggcc 660
Db 13308 AAGCAGGAATTGACGGGATTCACGATACAGCGTCTGCTGACCCCGGAGTTGCGGCC 13367
QY 661 accttgaagcgtgctagcacaactggcccccggcgcgacaaaccccgacgaccac 720
Db 13368 ACCATCGAGGGGGTGTGGCCAAACTGGCCGACCGGGGGTGCACCCCGATGACCCAG 13427
QY 721 accccggtcatcgacacacccccgatgcgccgcatcgaccgacaccccgagccaa 780
Db 13428 ACCCGGTCGTGGATGACACACCGGATGCGGACGCGGTGCCCGCGACACCCCGAGCCAA 13487
QY 781 gcccaacgcaaacacgacgagcgtcctggccgggtgcgcgctgagctccggaaaa 840
Db 13488 GCCCAACGACACCATGACGGTTTACTGGCGGGGTGCGCGGGTTCTGGCCTCCGGTCAG 13547
QY 841 ctgggcccaacaacggtcttcgcgtctcgtatcgtgtcacccacacccctgacccgacctg 900
Db 13548 CTAGGGGAGCATCGGGGTTCCCGGTGACCGTCTGTGTGGACACACCGCTTAAAGAGCTG 13607
QY 901 caaacgcgcgccgaagggttcaccgycggcgcgaccctgtacctatggccgcatgtg 960
Db 13608 GAAGCGCCACCGCAAGGGGTAAACCGTGTGGTTCGCGGGTCCCGATCTCGGACCTT 13667
QY 961 atccgcatgaccagccacgcccacactactcccccgcaagcgggaggtaccccccaagcg 1020
Db 13668 ATCCGGATGGCGAGCAAGCGCACCACTA-----TCTGGCA 13703
QY 1021 atcttcgaccggcacacccctggcgctgtatcacacaaacgcttagcclccccgccc 1080
Db 13704 TGTGTTGACGGCGTTAAGCCGTTGGCATTTGATCACACCAAGCGGTTAGCTTCCCGGCG 13763
QY 1081 cagcgatcatgtctgttcgcaacgacccgctgacccaaaccccgctgtgacgacccg 1140
Db 13764 CAGCGAATCATGTTGTACGCCAAGGATCGTGGCTGCTCCAGGCGGGTTGCGACGCCCG 13823
QY 1141 gcctaccacgccaagcccacacgctcaccgcttgagccagcagcgagcgtgagatc 1200
Db 13824 GCCTACCAACAGTGGGTCCACACGTAAAGCGCTGACACACCCCAACGTAACGATC 13883
QY 1201 accgagctgacctggcctgcgcccgcgaacacgactcgccgaaaaaggtgagaccacc 1260
Db 13884 AACGACCTCAGCTGGCTGGCGCCCGCCGACAAATCGCCTTTCGAAAAAGGCTGGAAACC 13943
QY 1261 cacaacaaccccgcgccacacggaatggctaccacacccccacctcgaccacggccaa 1320
Db 13944 CGCAAGAAGCGCAAGGGGACACTGAATGGCTACCGCGGCCCGCCACTTGGACCATGGCCAA 14003
QY 1321 ccccgaccacaacacttcacaccccggaagattctctcaacacgaagacgacgac 1380
Db 14004 CCACGCATCAATCGATACCAACCCCGGAGAAATCTCTGCGAACCGGACGACGACAA 14063
QY 1381 aaac 1384
Db 14064 CCAC 14067

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 20:15:40 ; Search time 366.19 Seconds
(without alignments)
3261.296 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcaa.....cgacgacaaacccgattgac 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1202.4	86.3	12412	17 AAT33537	BCG deletion regio
2	455	32.7	1535	16 AAQ89200	Mycobacterium tube
3	455	32.7	1535	17 AAT33656	M. tuberculosis ce
4	455	32.7	1535	19 AAV18647	DNA for M. tubercu
5	396	28.4	650	17 AAT33658	M. tuberculosis ma
6	396	28.4	650	19 AAV18649	DNA for M. tubercu
7	94.4	6.8	30690	21 AAA92301	S. avermitilis ave
8	89.8	6.4	1337	20 AA217263	Human gene express
9	89.4	6.4	12381	21 AA258381	Streptomyces averm
10	87.4	6.3	4466	21 AAA14663	Nucleotide sequenc
11	87.4	6.3	4478	21 AAA14661	Nucleotide sequenc

12	87.4	6.3	4547	21 AAA14664	Nucleotide sequenc
13	87.4	6.3	4571	21 AAA14662	Nucleotide sequenc
14	87.4	6.3	77536	21 AAA14651	Nucleotide sequenc
15	86.6	6.2	31422	21 AAA92302	S. avermitilis ave
c 16	82.6	5.9	77536	21 AAA14651	Nucleotide sequenc
c 17	82.6	5.9	114955	20 AAX53491	Human adenosine A1
c 18	82	5.9	58857	21 AAA58471	Nucleotide sequenc
c 19	81	5.8	1000	21 AAA02484	Human colon cancer
c 20	77.2	5.5	1218	21 AAA02488	Human colon cancer
21	75.6	5.4	4849	22 AAF25795	S. chrysomalinus ac
22	72	5.2	1359	20 AA217254	Human gene express
23	71.6	5.1	1635	22 AA122780	Probe #12713 for g
24	71.6	5.1	1635	22 AA148082	Probe #16768 used
25	71.6	5.1	1635	22 AA108454	Probe #8445 used t
26	71.6	5.1	1973	22 AA113556	Probe #3489 for ge
27	71.6	5.1	1973	22 AA134918	Probe #3604 used t
28	71.6	5.1	1973	22 AA101446	Probe #3437 used t
29	70.4	5.1	1401	22 AAH68136	C glutamicum codin
30	70.4	5.1	309400	22 AAH68534	C glutamicum codin
31	70.2	5.0	15720	22 AA129504	C899p determined c
32	70	5.0	925	20 AAV99229	DNA encoding an ac
33	70	5.0	925	21 AAA72501	LigAT2 domain from
c 34	69.2	5.0	1127	21 AAA02477	Human colon cancer
c 35	69	5.0	1459	21 AAA02528	Human colon cancer
36	69	5.0	4674	21 AAA14666	Nucleotide sequenc
37	69	5.0	4725	21 AAA14665	Nucleotide sequenc
38	69	5.0	4737	21 AAA14668	Nucleotide sequenc
39	69	5.0	4767	21 AAA14667	Nucleotide sequenc
40	69	5.0	4818	21 AAA14669	Nucleotide sequenc
c 41	68.8	4.9	1593	21 AAA02504	Human colon cancer
c 42	68.2	4.9	1075	22 AA121781	Probe #11714 for g
c 43	68.2	4.9	1075	22 AA147064	Probe #15750 used
c 44	68.2	4.9	1075	22 AA107466	Probe #7457 used t
c 45	68.2	4.9	1403	22 AA112593	Probe #2526 for ge

ALIGNMENTS

RESULT 1
AAT33537
ID AAT33537 standard; DNA: 12412 BP.
XX
XX AAT33537;
XX
DT 15-FEB-1998 (first entry)
XX
DE BCG deletion region 3 and flanking sequences.
XX
KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
XX
OS Mycobacterium bovis strain BCG.
XX
FH Key Location/Qualifiers
FT misc_feature 1406..10673
FT /*tag=a
FT /*note="BCG delta 1 deletion region"
XX
XX
PN WO9625519-A1.
XX
PD 22-AUG-1996.
XX
PF 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahairas GG, Slover CK;
XX
DR WPI: 1996-393419/39.
XX

Db	1067	acgagcatgcaaggtcaagtggctactcctgaccccccaagcgcgccacatttgaagcgg	1126
Qy	674	tgttagccaaactggcgcgcccgcgcgacaaaccccgacgacacaccccggtcatcg	733
Db	1127	tgttagccaaactggcgcgcccgcgcgacaaaccccgacgacacaccccggtcatcy	1186
Qy	734	acaccacccccgatggcgcgccatcgaccgacgacacccgacgacccaagccacgcaacc	793
Db	1187	acaccacccccgatggcgcgccatcgaccgacgacacccgacgacccaagccacgcaacc	1246
Qy	794	acgacgggctgctggcgcggtcgcgcgctgac	828
Db	1247	acgacgggctgctggcgcggtcgcgcgctgac	1281
RESULT 4			
ID	AAV18647		
XX	AAV18647	standard; DNA; 1535 BP.	
AC	AAV18647;		
XX	XX		
DT	03-JUL-1998	(first entry)	
XX	XX		
DE	DNA	for M. tuberculosis cellular uptake protein fragment.	
XX	XX		
KW	Cellular uptake protein; vaccine; infection; ds.		
XX	XX		
OS	Mycobacterium tuberculosis.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	1..1534	
FT	FT	/*tag=-,a^	
FT	FT	/note= "stop codon not given"	
XX	XX		
PN	WO9805784-A1.		
PD	12-FEB-1998.		
XX	XX		
PF	06-AUG-1997;	97WO-US13056.	
XX	XX		
PR	07-AUG-1996;	96US-0689411.	
XX	XX		
PA	(CONN-) CONNAUGHT LAB LTD.		
PA	(CORR) CORNELL RES FOUND INC.		
XX	XX		
PI	Chong P, Riley LW;		
DR	WPI; 1998-145620/13.		
DR	P-PSDB; AAW47541.		
XX	XX		
PT	Mycobacterium tuberculosis DNA - confers ability to enhance uptake of		
PT	therapeutic agents e.g. antibiotics, also useful in vaccines		
XX	XX		
PS	Disclosure; Pages 9-10; 82pp; English.		
XX	XX		
CC	The present sequence encodes a Mycobacterium tuberculosis cellular		
CC	uptake protein fragment, which confers on M. tuberculosis an		
CC	ability to enter mammalian cells and to survive within macrophages.		
CC	The protein can be used in a vaccine to prevent M. tuberculosis		
CC	infection, and provide for the uptake in cells of, e.g.		
CC	antibiotics, DNA fragments or anti-neoplastic agents. Antibodies		
CC	raised against it can be used to treat mammals already exposed to		
CC	M. tuberculosis, to induce a passive immunity and prevent disease		
XX	XX		
SQ	Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;		

```
Query Match      32.7%; Score 455; DB 19; Length 1535;
Best Local Similarity 100.0%; Pred. No. 4,1e-75;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

[illegible]

A DNA molecule (AAT33658) codes for a protein (AAW02303) that confers on *Mycobacterium tuberculosis* an ability to survive within macrophages. It is a fragment of a larger molecule (AAT33656) obt'd. by ligating *M. tuberculosis* genomic DNA fragments into pBluescript


```

XX 02-OCT-1998; 98US-0102748.
PR 11-MAR-1999; 99US-0123810.
PR 17-JUN-1999; 99US-0139650.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Reeves C, Chu D, Khosla C, Santi D, Wu K;
PI
XX WPI; 2000-317716/27.
DR
XX P-PSDB; AAY84727.
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy.
XX
XX Example 1; Page 82-85; 126pp; English.
XX
XX The present sequence represents an NheII-XhoII fragment that encodes
CC module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
CC endogenous acyltransferase (AT) domain replaced by the AT domain of
CC module 12 of the rapamycin PKS (which is specific for malonyl CoA).
CC FK-506 is a potent immunosuppressant, and acts through initial formation
CC of an intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,
CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
XX Sequence 4466 BP; 634 A; 1796 C; 1433 G; 603 T; 0 other;

```

```

Query Match      6.3%; Score 87.4; DB 21; Length 4466;
Best Local Similarity 47.2%; Pred. No. 9.5e-08;
Matches 265; Conservative 0; Mismatches 296; Indels 0; Gaps 0;
QY 828 cgctccgggaactgggccaacacaggtcttcctcgtctcgtggtcaccacac 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3545 cgaccccgaggacataccacacccgcgcacacacccgcgcacacccgcgc 3604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 888 cctgacccgacctgaaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3605 gcaacacacacctcaccacacgcgcacacacctcctcgtccacacacccgc 3664
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 948 catggccggtgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1007
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3665 cggcgccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3724
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1008 gtaccccccaggcgatcttgcacacgcgcgcgcgcgcgcgcgcgcgcgcgc 1067
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3725 cctcatcgaaacgcgcacaccccccacaccccccctcctcctcgcacaccc 3784
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1068 agcctcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3785 ccaccccccacccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1128 ctgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3845 caccacaccccccacacacaccccccctcacccccgaacacgcgcgcgcgcgc 3904
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1188 acgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 3905 cggctccggcaccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3964
QY 1248 aggtgacacacacacacacacacacacacacacacacacacacacacacac 1307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3965 cctcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4024
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1308 cgacacacacacacacacacacacacacacacacacacacacacacacac 1367
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4025 cgacacacacacacacacacacacacacacacacacacacacacacacac 14084
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1368 agacacacacacacacacacacacacacacacacacacacacacacacac 1388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4085 ccacacacacacacacacacacacacacacacacacacacacacacacac 4105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX RESULT 11
XX AAA14661
XX ID AAA14661 standard; DNA; 4478 bp.
XX
XX AC AAA14661;
XX
XX DT 08-AUG-2000 (first entry)
XX
XX DE Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
XX KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
XX OS Synthetic.
XX
XX ST Streptomyces hygroscopicus.
XX
XX FH Key Location/Qualifiers
XX CDS 3..4466
XX FT /*tag= a
XX FT /note= "no termination codon given"
XX
XX PN WO200020601-A2.
XX
XX PD 13-APR-2000.
XX
XX PF 01-OCT-1999; 99US-0122886.
XX
XX PR 02-OCT-1998; 98US-0102748.
XX
XX PR 11-MAR-1999; 99US-0123810.
XX
XX PR 17-JUN-1999; 99US-0139650.
XX
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX
XX DR WPI; 2000-317716/27.
XX
XX DR P-PSDB; AAY84725.
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
XX useful for treating e.g. transplant rejection, uveitis, multiple
XX sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
XX peripheral neuropathy.
XX
XX Example 1; Page 76-79; 126pp; English.
XX
XX The present sequence represents an AvrII-XhoII fragment that encodes
XX module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX endogenous acyltransferase (AT) domain replaced by the AT domain of
XX module 12 of the rapamycin PKS (which is specific for malonyl units).
XX FK-506 is a potent immunosuppressant, and acts through initial formation
XX of an intermediate complex with protein immunophilins known as FK-506
XX binding proteins. The nucleic acids are used for producing polyketide

```

compounds. The polyketide compounds can be used as immunosuppressants to prevent or treat transplant rejection, graft-versus-host disease or uveitis. They can also be used for treating e.g. alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, or scleroderma. They also have neurotrophic activity and can be used to promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. They can also be used for treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or peripheral neuropathies. They can also be used in agricultural and veterinary applications.

Query Match	6.3%;	Score 87.4;	DB 21;	Length 4478;
Best Local Similarity	47.2%;	Pred. No. 9.5e-08;		
Matches 265: Conservative	0;	Mismatches 296;	Indels 0;	Gaps 0;

[illegible]

RESULT 12
AAAL14664
ID AAAL14664 standard; DNA: 4547 BP.
XX
XX
XX AAAL14664;
XX
XX 08-AUG-2000 (first entry)
XX
XX Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
DE
DE FK-520: polyketide synthase; PKS: gene cluster; immunosuppressant;
XX
XX

immunophilin; FK-506 binding protein; polyketide synthase	KKW
transplant rejection; graft-versus-host disease; allogeneic	KKW
autoimmune chronic active hepatitis; inflammatory bowel disease	KKW
multiple sclerosis; primary biliary cirrhosis; scleroderma	KKW
neurite outgrowth; nerve regrowth; Parkinson's disease	KKW
Alzheimer's disease; stroke; traumatic spinal cord; peripheral neuropathy; ss	KKW
Synthetic.	XX
Streptomyces hygroscopicus.	OS
	OS
	XX
Key	location/Qualifiers
CDS	9..4535
FT	/*Tag= a
FT	/note= "no termination codon given"

WC200020601-A2.

13-APR-2000.

01-OCT-1999; 99WO-US22886.

02-OCT-1998; 98US-0102748.

11-MAR-1999; 99US-0123810.
17-JUN-1999; 99US-0139650.

(KOSA-) KOSAN BIOSCIENCES INC.

Reeves C, Chu D, Khosla C,

WPI; 2000-317716/27.

P-PSDB; AAY84728.

New isolated poly
useful for treati

sclerosis, Alzheimer peripheral neuropathy

Example 1; Page 85 88; 126pp; English.

The present sequence represents an NheII-XhoII fragment that encodes module 8 of the FK-520 polypeptide synthase (PKS) gene cluster with the endogenous aryltransferase (AT) domain replaced by the AT domain of module 13 of the rapamycin (PKS) which is specific for methylmalonyl CoA). FK-506 is a potent immunosuppressant, and acts through initial formation of an intermediate complex with protein immunophilins known as FK-506 binding proteins. The nucleic acids are used for producing polypeptide compounds. The polypeptide compounds can be used as immunosuppressants to prevent or treat transplant rejection, graft-versus-host disease or arthritis. They can also be used for treating e.g. alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, or scleroderma. They also have neurotrophic activity and can be used to promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. They can also be used for treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or peripheral neuropathies. They can also be used in agricultural and veterinary applications.

Sequence 4547 BP; 658 A; 1851 C; 1437 G; 601 T; 0 other;

Query Match	6.3%	Score 87.4;	DB 21;	Length 4547;
Best Local Similarity	47.2%	Pred. No. 9.5e-08;		
Matches 265; Conservative	0;	Mismatches 296;	Indels 0;	

Qy	828	cgctccgggaac:gggccaacaaagtgcttccgctctcgatcgtggcaccac:	887
Db	3626	cgaccctgaagacataccacccggccacaccgcgctctgaccgcct:	3685
Qy	888	cctgaccacctgcacacggcgccggcaagggttcacggcgggcaccctgctac:	947

QY	1308	cgaccacgcccgaacccgcaccaacaacttcacaccacccccgaagattcctcccaacca	1367
Db	4130	cgacccccaactgcaccacctcacccacatcccccaacccctcacccgcaictt	4189
QY	1368	agacgacgacgacaaaaccga	1388
Db	4190	ccataccgcccaccctcga	4210
RESULT	14		
ID	AAA14651		
XX	AA14651 standard; DNA; 77536 BP.		
AC	AAA14651;		
DT	08-AUG-2000 (first entry)		
DE	Nucleotide sequence of the FK-520 biosynthetic gene cluster.		
KW	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;		
KW	Streptomyces hygroscopicus var. ascomyceticus; immunophilin;		
KW	FK-506 binding protein; polyketide compound; transplant rejection;		
KW	graft-versus-host disease; uveitis; alopecia universalis;		
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;		
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;		
KW	neurile outgrowth; nerve regrowth; Parkinson's disease;		
KW	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;		
XX	peripheral neuropathy; ss.		
OS	Streptomyces hygroscopicus.		
Key	Location/Qualifiers		
FT CDS	complement (412..1836)		
FT	/*tag= a		
FT	/note= "fkbW gene"		
FT CDS	complement (2020...3579)		
FT	/*tag= b		
FT	/note= "fkbV gene"		
FT CDS	3969..4496		
FT	/*tag= c		
FT	/note= "fkbR2 gene"		
FT CDS	complement (4595..5488)		
FT	/*tag= d		
FT	/note= "fkbR1 gene"		
FT CDS	5601...6818		
FT	/*tag= e		
FT	/note= "fkbE gene"		
FT CDS	6808..8052		
FT	/*tag= f		
FT	/note= "fkbF gene"		
FT CDS	8156..8824		
FT	/*tag= g		
FT	/note= "fkbG gene"		
CDS	complement (9122..9883)		
FT	/*tag= h		
FT	/note= "fkbH gene"		
FT CDS	complement (9894..10994)		
FT	/*tag= i		
FT	/note= "fkbI gene"		
FT CDS	complement (10987...11247)		
FT	/*tag= j		
FT	/note= "fkbJ gene"		
FT CDS	complement (11244...12092)		
FT	/*tag= k		
FT	/note= "fkbK gene"		
FT CDS	complement (12113...13150)		
FT	/*tag= l		
FT	/note= "fkbL gene"		
FT CDS	complement (13212...23988)		
FT	/*tag= m		
FT	/note= "fkbC gene"		
FT misc_feature	complement (13452...13662)		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 20:24:00 ; Search time 165.4 Seconds
(without alignments)
1907.399 Million cell updates/sec

Title: US-09-785-904-1
Perfect score: 1393
Sequence: 1 tgggttcgggtagccggaa.....cqcacacaaacccqattqac 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

```
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
c	1	1301.4	93.4	4411529	4	US-09-103-840A-1
	2	1202.4	86.3	12412	1	US-08-390-878-18
	3	455	32.7	1535	3	US-08-464-052-1
	4	455	32.7	1535	4	US-08-461-002-1
	5	455	32.7	1535	4	US-08-689-411-1
	6	455	32.7	1535	5	PCR-US94-09863-1
	7	396	28.4	650	3	US-08-464-052-5
	8	396	28.4	650	4	US-08-461-002-5
	9	396	28.4	650	4	US-08-689-411-5
	10	101.2	7.3	4403765	4	US-09-103-840A-2
c	11	101.2	7.3	4411529	4	US-09-103-840A-1
	12	78.4	5.6	4403765	4	US-09-103-840A-2
	13	70.8	5.1	1931	2	US-09-130-114-2
	14	70	5.0	925	3	US-08-858-003-1
	15	70	5.0	925	3	US-09-078-166-1
	16	70	5.0	925	4	US-08-997-467-1
	17	69.2	5.0	1722	4	US-09-385-028-15
	18	69.2	5.0	11604	4	US-09-385-028-13
	19	69.2	5.0	15079	4	US-09-385-028-1
	20	64.8	4.7	1248	4	US-09-105-537-7
c	21	64.8	4.7	13613	4	US-09-105-537-3
	22	62.8	4.5	44377	4	US-08-804-227C-7
	23	62.8	4.5	44377	2	US-08-804-198-1
	24	60.6	4.4	1280	4	US-09-060-756-4
	25	60.2	4.3	2712	3	US-09-025-691-4
	26	59.6	4.3	1294	3	US-09-025-691-2
	27	59.6	4.3	1590	4	US-09-434-288-1

28	57.8	4.1	3468	1	US-07-951-715A-2	Sequence 2, Appl 1
29	57.8	4.1	3458	2	US-08-459-448A-2	Sequence 2, Appl 1
30	57.8	4.1	3468	3	US-08-459-595A-2	Sequence 2, Appl 1
31	57.8	4.1	3468	3	US-08-459-504B-2	Sequence 2, Appl 1
32	57.8	4.1	3468	3	US-08-459-44A-2	Sequence 2, Appl 1
33	57.8	4.1	3468	3	US-09-053-549-3	Sequence 3, Appl 1
34	57.2	4.1	30001	1	US-08-125-468-1	Sequence 1, Appl 1
35	57.2	4.1	30001	2	US-08-474-933-1	Sequence 1, Appl 1
36	57	4.1	1681	4	US-09-434-288-7	Sequence 7, Appl 1
37	56.6	4.1	1155	2	US-08-387-942C-18	Sequence 18, Appl 1
38	56.6	4.1	12588	2	US-08-387-942C-1	Sequence 1, Appl 1
39	56.2	4.0	1947	1	US-07-951-715A-3	Sequence 3, Appl 1
40	56.2	4.0	1947	2	US-08-459-448A-3	Sequence 3, Appl 1
41	56.2	4.0	1947	3	US-08-459-595A-3	Sequence 3, Appl 1
42	56.2	4.0	1947	3	US-08-459-504B-3	Sequence 3, Appl 1
43	56.2	4.0	1947	3	US-08-459-44-3	Sequence 3, Appl 1
44	56.2	4.0	3468	1	US-07-951-715A-4	Sequence 4, Appl 1
45	56.2	4.0	3468	1	US-07-951-715A-8	Sequence 8, Appl 1

ALIGNMENTS

```

RESULT      1
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

[illegible]

[illegible]

```

RESULT      3
US-08-464-052-1
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof

```

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,052
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ. IN NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-464-052-1

[illegible]

RESULT 4
US-08-461-002-1

QY	553	gacctcacccagacaccgaacgcgcccgcaaaacgcggcatcacccttgagcaaacaggcaatac	612
DB	121	GACCTCACCCAGACACCGAACGCGGCCGCAAAACGGGCATCACCCTGAGCAACACGCAATATC	180
QY	613	gacggcatgtcaacgggttaagtggctactctgaccccccaagcgggcccacacctttgaagcc	672
DB	181	GACGGCATGTCAACGGTCTAAGTTGGCTACTCTACCCCCCAAGCGGGCCACCTTTGAAGCC	240
QY	673	gtctagtcaccaaatgucgcgcgccttgygcgacaacaacccagacagacacaccccggtcatac	732
DB	241	GTGCTAGTCACCAAACTGGCGCCGCCCGCGCGGACCAACCCCGACACACACACCCCGTTCATC	300
QY	733	gacaccacccccatgcyqgcgcgcattgcacgcgcacacccgcagcgaagcccaacgcgaac	792
DB	301	GACACCACCCCGCATGCGCGCCGCATCGACCGGAGACACCCGACGACGCCCAALGCAAC	360
QY	793	cacjacgggtgtctgycgcgggtgcgcgcgcgtgatc	828
DB	361	CACGACGGGTGTGTCGCGGGCTGCGGCGTGATC	396
 RESULT 9			
US-08-689-411-5			
; Sequence 5, Application US/08689411			
; Patent No. 6224881			
; GENERAL INFORMATION:			
; APPLICANT: Riley M.D., Lee W.			
; APPLICANT: Chong, Pele			
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR			
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES			
; NUMBER OF SEQUENCES: 14			
; THEREOF OF INVENTION: THEREOF			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP			
; STREET: Clinton Square, P.O. Box 1051			
; CITY: Rochester			
; STATE: New York			
; COUNTRY: U.S.A.			
; ZIP: 14603			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; FILING DATE: US/08/689,411			
; CLASSIFICATION: 546			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Goldman, Michael L.,			
; REGISTRATION NUMBER: 30,727			
; REFERENCE/HOCKET NUMBER: 19603/187			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (716) 263-1304			
; TELEFAX: (716) 263-1600			
; INFORMATION FOR SEQ ID NO: 5:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 650 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: unknown			
; MOLECULE TYPE: DNA (genomic)			
US-08-689-411-5			
Query Match 28.4%; Score 396; DB 4; Length 650;			
Best Local Similarity 100.0%; Prod. No. 1e-75;			
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	433	gtggatgtgtccacrcgcagccagccgcgaagccgacctggccggccaagcgctcaatal	492
DB	1	GTGATGTGTCACCCCGGCGAGCGCCGCAAGCCGACCTGGCGCGCAAGCGCTCAATAT	60


```
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          5.6%; Score 78.4; DB 4; Length 4403765;
Best Local Similarity 97.2%; Pred. No. 8.9e-08;
Matches 1344; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 7 cgggtagccgcgaacgattgtcgaagtttttgatcgcgtggtatgcgcgagctggacgcgt 66
D 1780705 CGAATAGCCGGAGGAGCTTGTGAGGCTTTTGATGCGCTGGATGCCGAGCTGGACCGCT 1780646

QY 67 tgaacaggtgtcttttgaggtgtttgaacacccccagaacggtggtctctggaacgtc 126
D 1780645 TGAACAGGTGTCTTTTGAGGTGTTGACACCCCGGAACGGCTGCGGTCTCTGGAACGTC 1780586

QY 127 tgaatcttggtgcgcggtaccggcggtlgggtcacgctgtgatcaacacacttgacg 186
D 1780585 TGAATGCTTGTGTCGCCGCTACCGCGGTGGGCGACACGCTTGATCAACCACTCGACA 1780526

QY 187 ccaacgcagcgaaggaactggcggaactgtgctgctgctgctgctgctgctgctgct 246
D 1780525 CCAACGCCAGCGAGGAAGAACTGGGCGCGACGCTGTGTCGCGCTGGCCAAACCGTTAC 1780466

QY 247 gcatcaccagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
D 1780465 GCATCAACCAAGCCGACCGCGCCCTACGCATCGCGGACCGCGCGCGATCTCGACCTCGTC 1780406

QY 307 gacactcaccggtgaacgcgttagccccacagltgacgcgcacccgcgcgcgcgcgcgc 366
D 1780405 GAGCACTCACCGGGAACCGCTAGCCCCACACAGTTGACCGCGCACCGCCCAACGCC 1780346

QY 367 agggcctgacgcgagggcagcagtcacaaagtattcgccctttttgcccacactgccc 426
D 1780345 AGGGCCTGATCGCGAGGCGGCACATCAAGATGATTCGGCGCCTTTTTCGCCACCTGCC 1780286

QY 427 gccgcggttgatgtgtccacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 486
D 1780285 GCCGCGGTGGATGTGTCCACCCGCGCAGCGCGCGGAAACCGACCTGGCGCGCAAGCGCT 1780226

QY 487 caatatcgtccgcagagctggccgcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 546
D 1780225 CAATATCTCCGACGAGCTGGCCCCGCTACGCCCGAGGGGTATGGACTGGCTACACTTC 1780166

QY 547 gacgcgcactcacgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 606
D 1780165 GACGGCGACCTCACGACACCGACCGCGCGCGCGGAAACCGGCGATCACCTGACACCCAG 1780106

QY 607 caatacgcgcgcgtacgcgcgttaagtgtacctgaccccccccaagcgcgcgcgcacctt 666
D 1780105 CAATACGACGCATGTACAGGCTAAAGTGGCTAGCTGACCCCGCAAGCGCGGCACCTTT 1780046

QY 667 gaagcgtgtgtagcaaaactggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 726
D 1780045 GAACCGTGTAGCCAAACTGGCGCGCGCGCGCGCGGCGGCAACCCCGACACACCCCG 1779986

QY 727 gtcacgcacacccccgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 786
D 1779985 GTCATGACACACACCCCGATGGCGCGCATCGACCGGACACCGCGGCGCGCGCGCGCG 1779926

QY 787 cgaacacacgcgcgcgtactgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 846
D 1779925 CGCAACACACGCGGCTGTGCGCGCGGCGCGCGCGCTGATCGCTCCGGGAAACTGGGC 1779866

QY 847 caacacacggtcttccgctctgctgctgctgctgctgctgctgctgctgctgctgctgct 906
D 1779865 CAACACACGCTTCTCCGCTCTGATGCTGGTCAACCAACCAACCGCTGACCGACCTGCA 1779806
```

```
QY 907 ggccgcgcgaagggcttcaaccggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 966
D 1779805 GGCCGCGCAAGGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779746

QY 967 atgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1026
D 1779745 ATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1779686

QY 1027 gaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1086
D 1779685 GACCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779626

QY 1087 atcatgctgtlccgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1146
D 1779625 ATCATGCTGTTCGCAACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 1779566

QY 1147 cacagccaaqcccccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1206
D 1779565 CACAGCCAAAGCCACCGCTACCGCGCTGGACGACCGCGCGCGCGCGCGCGCGCGCGCG 1779506

QY 1207 ctgacccclggcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1266
D 1779505 CTCACCCCTGGCGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779446

QY 1267 aacacccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1326
D 1779445 AACACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779386

QY 1327 accaaacacttccaccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1386
D 1779385 ACCAAACACTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779326

QY 1387 gat 1389
D 1779325 GAT 1779323

RESULT 13
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Epithomes
; FILE REFERENCE: 0867/10903051
; CURRENT APPLICATION NUMBER: US/09/140,114
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match          5.1%; Score 70.8; DB 2; Length 1931;
Best Local Similarity 43.1%; Pred. No. 8.9e-07;
Matches 342; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

QY 578 gcaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
D 258 gccacctgtctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgt 317

QY 638 acctgaccccccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 697
D 318 tctctcccgctcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcct 377

QY 698 gcgcgaccacccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 757
```


[illegible]

RESULT 14
US-08-858-003-1

; Sequence 1, Application US/08858003

; Patent No. 6060234

; GENERAL INFORMATION:

APPLICANT: Katz, Leonard

APPLICANT: Stassi, Diane L.

APPLICANT: Summers Jr., Richard G.
 APPLICANT: Summers Jr., Richard G.

APPLICANT: SUMMERS JR., RICHARD G.
APPLICANT: Ruan, Xiaolan

APPLICANT: KUBIL, XIROAN
; APPLICANT: Pereda-Lopez, Ana

APPLICANT: PEREUD-LOPEZ, ANA
APPLICANT: KAKAVAS, STEPHAN J.

APPLICANT: Kakavas, Stephan J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES

FILE OF INVENTION:	NOVEL POLYKETIDE DERIVATIVES
TITLE OF INVENTION:	AND RECOMBINANT METHODS FOR MAKING SAME
FILE OF INVENTION:	NOVEL POLYKETIDE DERIVATIVES
TITLE OF INVENTION:	AND RECOMBINANT METHODS FOR MAKING SAME

;	TITLE OF INVENTION:	AND
;	NUMBER OF SEQUENCES:	34

NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Rd.

CITY: Abbott Park

STATE: Illinois

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

```

; OPERATING SYSTEM: DOS

```

```

; SOFTWARE: FastSEQ version 2.0

```

1 CURRENT APPLICATION DATA:
 2 APPLICATION NUMBER: US/08/858,003
 3 FILING DATE: 16-MAY-1979
 4 CLASSIFICATION: 435
 5 PRIOR APPLICATION DATA:
 6 APPLICATION NUMBER:
 7 FILING DATE:
 8 ATTORNEY/AGENT INFORMATION:
 9 NAME: Dianne Casuto
 10 REGISTRATION NUMBER: P-40,943
 11 REFERENCE/DOCKET NUMBER: 4952.US.P2
 12 TELECOMMUNICATION INFORMATION:
 13 TELEPHONE: (847)-938-3137
 14 TELEFAX: (847)-938-2623
 15 TELEX:
 16 INFORMATION FOR SEQ. ID. NO.: 1:
 17 SEQUENCE CHARACTERISTICS:
 18 LENGTH: 925 base pairs
 19 TYPE: nucleic acid
 20 STRANDEDNESS: double
 21 TOPOLOGY: linear
 22 US-08-858-003-1

Query Match 5.0%; Score 70; DB 3; Length 925;

[illegible]

RESULT 15
US-09-078-166-1
; Sequence 1, Application US/09078166
; Patent No. 6063561
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 18:47:35 ; Search time 3835.7 Seconds
(without alignments)
3902.513 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcgaa.....cgacgacaaacccgattgac 1393

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estlir:*

4: em_estcom:*

5: em_estpi:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_hic:*

10: gb_estl:*

11: gb_est2:*

12: gb_hic:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_fod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	105.2	7.6	1197	11	BI416470 hasp001xa
C 2	104.6	7.5	1223	11	BI416573 hasp001xg
C 3	100.4	7.2	1015	11	BI416650 hasp001xm
C 4	99.2	7.1	924	10	BE455162 HVSMEh009
C 5	98.4	7.1	1224	11	BI416537 hasp001xe
C 6	98	7.0	931	11	BI416575 hasp001xg
C 7	97.2	7.0	1159	11	BF260036 HVSMEf002
C 8	96.6	6.9	289	10	AI864419 w152f10.x
C 9	95.8	6.9	863	11	BI416699 hasp001xp
C 10	95.8	6.9	1348	11	BF628951 HVSMEb000
C 11	95.6	6.9	1016	13	AL249930 Tetraodon
C 12	95.4	6.8	825	11	BG786255 SEAUWC006

C 13	95	6.8	918	13	AQ895329
C 14	95	6.8	966	13	AQ748706
C 15	94.4	6.8	782	11	BG809582
C 16	94.4	6.8	855	10	AL572700
C 17	94.4	6.8	1006	10	BE034822
C 18	94.4	6.8	1035	11	BF254665
C 19	94	6.7	873	11	BG786249
C 20	94	6.7	1192	11	BF256672
C 21	93.6	6.7	897	11	BI416517
C 22	93.6	6.7	1299	11	BF267681
C 23	93.2	6.7	1050	11	BF267685
C 24	93	6.7	952	11	BG441341
C 25	92.4	6.6	1088	11	BF256580
C 26	92	6.6	1195	11	BF256755
C 27	91.4	6.6	1039	11	BF259783
C 28	90.2	6.5	1042	11	BI416535
C 29	90	6.5	1280	11	BF864339
C 30	89.6	6.4	1166	11	BF256751
C 31	89.2	6.4	1186	11	BF256757
C 32	88.4	6.3	1325	11	BF256543
C 33	88.2	6.3	1122	5	BF628976
C 34	88	6.3	884	13	AQ782679
C 35	88	6.3	883	13	AZ186646
C 36	87.6	6.3	915	13	AZ189327
C 37	87.4	6.3	1086	11	AZ195437
C 38	87.2	6.3	1080	11	BG809658
C 39	87	6.2	834	11	BI416623
C 40	87	6.2	918	11	BI416657
C 41	87	6.2	1504	10	BE421615
C 42	87	6.2	806	11	BI416659
C 43	86.8	6.2	806	11	BI416659
C 44	86	6.2	1277	11	BG447302
C 45	85.8	6.2	893	13	AZ199082

ALIGNMENTS

RESULT 1
BI416470/C

LOCUS

DEFINITION

BI416470 1197 bp mRNA
hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Heterobasidion annosum/Pinus sylvestris mixed EST library cDNA clone hasp001xa01f, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

15-AUG-2001
Scots pine infection stage subtraction cDNA library (hasp) Heterobasidion annosum/Pinus sylvestris mixed EST library cDNA clone hasp001xa01f, mRNA
sequence.
BI416470
BI416470.1 GI:15187493
EST.
Heterobasidion annosum/Pinus sylvestris mixed EST library.
Eukaryota; mixed EST libraries.
1 (bases 1 to 1197)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, K.A.
Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
Location/Qualifiers
1..1197
/organism="Heterobasidion annosum/Pinus sylvestris mixed EST library"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"

```
/dev_stage="Seedling roots of scots pine were infected for
6 weeks with H. annosum"
notes="Vector: pT-Adv; Site_1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."

```

[illegible][illegible]

RESULT 2
 LOCUS B1416573/c
 DEFINITION
 B1416573 1223 bp mRNA
 hasp001xg1f Heterobasidion annosum - Scots pine infection stage
 subtraction cDNA library (hasp) Heterobasidion annosum/Pinus
 sylvestris mixed EST library cDNA clone hasp001xg1f, mRNA
 sequence.
 ACCESSION B1416573
 VERSION B1416573.1 GI:15187596
 KEYWORDS EST.
 SOURCE Heterobasidion annosum/Pinus sylvestris mixed EST library.
 ORGANISM Heterobasidion annosum/Pinus sylvestris mixed EST library.
 Eukaryota; mixed EST libraries.
 REFERENCE 1 (bases 1 to 1223)
 Asiegbu, F.O., Nahaikova, J., Choi, W., Stenlid, J. and Dean, R.A.
 Expressed sequence tags of randomly selected cDNA clones from the
 interaction of the root rot fungus (Heterobasidion annosum) with
 seedling roots of Scots pine (Pinus sylvestris)
 unpublished (2001)
 JOURNAL Contact: Fred O. Asiegbu
 COMMENT Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
 Sweden
 Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@mykopat.slu.se

FEATURES	Source
seq primer: 17 primer.	location/Qualifiers
	1..1223
	/organism "Heterobasidion annosum/Pinus sylvestris mixed EST library"
	/db_xref="taxon:169015"
	/clone "hshp01x111"
	/clone_lib "Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
	/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
	/note="Vector: pT-Adv; Site:1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FP5)."
BASE COUNT	49 a 24 c 1047 g 34 t 69 others
ORIGIN	

	Query Match	7-5%	Score 104.6;	DB 11;	Length 1223;
	Best Local Similarity	41.7%;	Pred. No.	4.1e-09;	Mismatches 0;
	Matches 469;	Conservative	Indels	Gaps	1;
QY	246	cgcataccaaagcggacgcccgcgatcggaagccccgccgtatcttcggaccttgt	305		
Dd	1162	CCCCCCCCCCCGCCGCCCCTCAGTGGGTTCATTTGCACCAACTTGGAACCTG	1103		
QY	306	caggaccactcaaccggtagaacccagttagccgccaacgccaacgcccacaacgc	365		

[illegible]

RESULT 3
BI416650/c

```

LOCUS       B1416650             1015 bp                15-AUG-2001
DEFINITION  hasp001xm01f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Heterobasidion annosum/Pinus
            sylvestris mixed EST library cDNA clone hasp001xm01f, mRNA
            sequence.
ACCESSION   B1416650
VERSION     B1416650.1  GI:15187673
KEYWORDS    EST.
SOURCE      Heterobasidion annosum/Pinus sylvestris mixed EST library.
            Eukaryote; mixed EST libraries.
ORGANISM    1 (bases 1 to 1015)
            Asiegbu, F.O., Nahaiakova, J., Choi, W., Stenlid, J. and Dean, R.A.
            Expressed sequence tags of randomly selected cDNA clones from the
            infection of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
            Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: "7 primer."
            Location/Qualifiers
                1..1015
                    /organism="Heterobasidion annosum/Pinus sylvestris mixed
                    EST library"
                    /db_xref="taxon:169015"
                    /clone_lib="hasp001xm01f"
                    /clone_lib="Heterobasidion annosum - Scots pine infection
                    stage subtraction cDNA library (hasp)"
                    /dev_stage="Seedling roots of scots pine were infected for
                    6 days with H. annosum"
                    /note="Vector: pT-Adv; Site:1: EcoRI; The subtractive
                    hybridization cDNA library was constructed from scots
                    pine roots infected for 6-days with mycelia of
                    Heterobasidion annosum (FP5)."
BASE COUNT  21 a          6 c  961 g  16 t          11 others
ORIGIN

```

[illegible]

KEYWORDS
SOURCE Heterobasidion annosum/Pinus sylvestris mixed EST library.
ORGANISM Heterobasidion annosum/Pinus sylvestris mixed EST library
REFERENCE 1 (bases 1 to 863)
AUTHORS Asiegbu F.O., Nahalkova,J., Choi,W., Stenlid J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES
source Location/Qualifiers
1..863
/organism="Heterobasidion annosum/Pinus sylvestris mixed
EST library"
/db_xref="taxon:169015"
/clone="hasp001xp01f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Aov; Site_1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."

	BASE COUNT	12 a	21-c	700 g	12 t	118 others
ORIGIN						

Query Match 6.9%; Score 95.8; DB 11; Length 863;
Best Local Similarity 43.1%; Pred. No. 1.4e-07;
Matches 292; Conservative 0; Mismatches 386; Indels 0; Gaps 0;

QY	689	cgcgcccgcgagcaaacccgacgacacaccgggtcatgacacaccccgcgatg	748
Db	838	CCCCCCCCCCCCCCCCCCNCCCNCCCNCNCCCCCCCCCCCCCCCCCCCCC	779
QY	749	cggcgccatgacgcgcgacacccgcgcgaagcccaacgcaacacgacgaggctgtg	808
Db	778	CCCCNNNCCCCCCCCNCCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC	719
QY	809	ccngcgtgcgcgctgatgcctccgggaactggqccaacacaacagattcttcgttct	868
Db	718	CCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNNNNNNCCCC	659
QY	869	cgatcgttggtcaccaaccctgacgcacctgcaaacggcgcggaagggtttcaccy	928
Db	658	CCNCCNCCNCCCC	599
QY	929	gcgcgcgacactgtaccactggccgatlgatccgcatgatccagcagcaccacact	988
Db	598	CCNCCNCCNCCCC	539
QY	989	actccccgaacgggagtacccccaggcgatcttcgaccacggcacaccctggcgc	1048
Db	538	CCNCCNCCNCCCC	479
QY	1049	tgtatcacacaaacgctagcctccccggcccgagggatcgtgttgtcgcaacgacc	1108
Db	478	CNCC	419
QY	1109	gcgcgtcacaaaaccggctgtgagcacccgcctcatacagcgaagcccaacgata	1168
Db	418	CCNCCNCCNCCCC	359
QY	1169	ccgcgtgacacgacgcgcgacgcgacatcacccagctgaccccttgctgcgcgcgc	1228

[illegible]

Search completed: April 2, 2002, 21:47:49
Job time: 10814 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:47:50 ; Search time 366.19 Seconds
(without alignments)
1060.565 Million cell updates/sec

Title: US-09-785-904-2

Perfect score: 453

Sequence: 1 gatcgcgagcgacatca.....cgggctgcgcgctgctc 453

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	408.4	90.2	1535	16 AAQ89200 Mycobacterium tube
2	408.4	90.2	1535	17 AAT33556 M. tuberculosis ce
3	408.4	90.2	1535	19 AAV18647 DNA for M. tubercu
4	366.8	81.0	12412	17 AAT33537 BCG deletion regio
5	351	77.5	650	17 AAT33658 M. tuberculosis ma
6	351	77.5	650	19 AAV18649 DNA for M. tubercu
7	57.2	12.6	30690	21 AAQ92301 S. avermitilis ave
8	55.6	12.3	12381	21 AAQ92301 S. avermitilis ave
9	55.6	12.3	31422	21 AAQ92302 S. avermitilis ave
10	52.4	11.6	3198	20 AAX02974 Human li-lra BAC c
11	49.4	10.9	77536	21 AAA14651 Nucleotide sequenc

12	47.8	10.6	2697	13 AAQ22957 Sequence of beta-1
13	47.2	10.4	1000	21 AAQ02484 Human colon cancer
14	47.2	10.4	1454	20 AAQ17038 Human gene express
15	46.6	10.3	1337	20 AAQ17263 Human gene express
16	45.8	10.1	114955	20 AAQ53491 Human adenosine A1
17	45.6	10.1	14466	21 AAA14661 Nucleotide sequenc
18	45.6	10.1	14478	21 AAA14661 Nucleotide sequenc
19	45.6	10.1	1457	21 AAA14661 Nucleotide sequenc
20	45.6	10.1	14571	21 AAA14662 Nucleotide sequenc
21	45.6	10.1	77536	21 AAA14651 Nucleotide sequenc
22	45.4	10.0	2888	17 AAT59268 Streptomyces prist
23	45	9.9	567	21 AAA29550 HIV codon altered
24	44.6	9.8	6085	18 AAT70153 S.Lonipisporoflavus
25	43.4	9.6	1028	13 AAQ27091 XTY26 probe. Homo
26	43.4	9.6	1028	22 AAQ91542 1.0kb PstI fragmen
27	43.4	9.6	1028	22 AAS01347 Human Fragile X Sy
28	43.2	9.5	1107	22 AAH74540 Nucleotide sequenc
29	43.2	9.5	2712	20 AAZ08825 Streptomyces albid
30	43.2	9.5	2712	22 AAH74538 Nucleotide sequenc
31	42.6	9.4	1018	16 AAT02326 Human cardiac hype
32	42.6	9.4	1203	20 AAZ00869 Human cardiotoxphi
33	42.6	9.4	1485	21 AAZ61590 Human c-Maf cDNA.
34	42.6	9.4	1485	21 AAZ61590 DNA encoding a hum
35	42.6	9.4	1539	20 AAX87267 DNA encoding a hum
36	42.6	9.4	1539	21 AAA99914 cDNA clone encodin
37	42.6	9.4	1539	21 AAA72705 cDNA encoding huma
38	42.6	9.4	1539	21 AAA72706 Human cardiotoxphi
39	42.6	9.4	1539	21 AAA72706 Human cardiotoxphi
40	42.6	9.4	1539	22 AAC90578 cDNA encoding nove
41	42.6	9.4	1539	22 AAC90578 Human PRO882 cDNA.
42	42.4	9.4	494	19 AAV64542 M. tuberculosis im
43	42.4	9.4	494	19 AAV44433 Mycobacterium tube
44	42.4	9.4	494	20 AAZ19343 M. tuberculosis an
45	42.4	9.4	494	20 AAZ19131 M. tuberculosis re

ALIGNMENTS

RESULT 1
AAQ89200
ID AAQ89200 standard; DNA; 1535 BP.
XX
AC AAQ89200;

DT 09-OCT-1995 (first entry)

DE Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
protein.

DE Vaccine; tuberculosis; ss.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT CDS 1..1535

FT /*Lag= a

PN WU9506726-A.

PD 09-MAR-1995.

XX 01-SEP-1994; 94WO-US09863.

XX 02-SEP-1993; 93US-0118442.

XX (CORR) CORNELL RES FOUND INC.

XX Riley LW;

XX WPI; 1995-115442/15.

XX P-PSDB; AAR71931.

XX

[illegible]

Matches	182;	Conservative	0;	Mismatches	217;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	------	--------	----	------	----

FTT

```
FT /note= "fkbf gene"
FT 8156.8824
CDS
FT /*lag= 9
FT /note= "fkbg gene"
FT
```

```

FT      CDS      complement (9122..9883)
FT      /tag= h
FT      /note= "fkbh gene"
FT      CDS      complement (9894..10994)
FT      /tag= i
FT      /note= "fkb1 gene"
FT      CDS      complement (10987..11247)
FT      /tag= j
FT      /note= "fkb1 gene"
FT      CDS      complement (11244..12092)
FT      /tag= k
FT      /note= "fkbk gene"
FT      CDS      complement (12113..13150)
FT      /tag= l
FT      /note= "fkbl gene"
FT      CDS      complement (13212..23988)
FT      /tag= m
FT      /note= "fkbc gene"
FT      CDS      complement (13452..13662)
FT      /tag= n
FT      /note= "ACPe"
FT      misc_feature complement (13761..14394)
FT      /tag= o
FT      /note= "KR6"
FT      misc_feature complement (14517..15294)
FT      /tag= p
FT      /note= "ER6"
FT      misc_feature complement (15438..16587)
FT      /tag= q
FT      /note= "dehydratase domain (DH) 6"
FT      misc_feature complement (16587..17820)
FT      /tag= r
FT      /note= "acylttransferase domain (AT) 6"
FT      misc_feature complement (17820..19053)
FT      /tag= s
FT      /note= "KS6"
FT      misc_feature complement (19116..19326)
FT      /tag= t
FT      /note= "ACPs"
FT      misc_feature complement (19464..20097)
FT      /tag= u
FT      /note= "KR5"
FT      misc_feature complement (20241..21420)
FT      /tag= v
FT      /note= "DH5"
FT      misc_feature complement (21420..22653)
FT      /tag= w
FT      /note= "AT5"
FT      misc_feature complement (22653..23892)
FT      /tag= x
FT      /note= "KS5"
FT      CDS      complement (23992..46573)
FT      /tag= y
FT      /note= "fkbB gene"
FT      misc_feature complement (24163..24373)
FT      /tag= z
FT      /note= "ACP4"
FT      misc_feature complement (24997..26146)
FT      /tag= aa
FT      /note= "DH4 (inactive)"
FT      misc_feature complement (26146..27430)
FT      /tag= ab
FT      /note= "AT4"
FT      misc_feature complement (27430..28684)
FT      /tag= ac
FT      /note= "KS4"
FT      misc_feature complement (28750..28960)
FT      /tag= ad
FT      /note= "ACP3"
FT      misc_feature complement (29092..29740)
FT      /tag= ae
FT      /note= "KR3"
FT      misc_feature complement (29869..31018)

```

```

FT      /tag= af
FT      /note= "DH3 (inactive)"
FT      misc_feature complement (31018..32185)
FT      /tag= ag
FT      /note= "AT3"
FT      misc_feature complement (32185..33439)
FT      /tag= ah
FT      /note= "KS3"
FT      misc_feature complement (33505..33715)
FT      /tag= ai
FT      /note= "ACp2"
FT      misc_feature complement (33823..34480)
FT      /tag= aj
FT      /note= "KR2"
FT      misc_feature complement (34606..35749)
FT      /tag= ak
FT      /note= "ACP1"
FT      misc_feature complement (35749..37144)
FT      /tag= al
FT      /note= "DH2 (inactive)"
FT      misc_feature complement (37145..38296)
FT      /tag= am
FT      /note= "AT2"
FT      misc_feature complement (37145..38296)
FT      /tag= an
FT      /note= "KS2"
FT      misc_feature complement (38371..38581)
FT      /tag= ao
FT      /note= "ACP1"
FT      misc_feature complement (38677..39307)
FT      /tag= ap
FT      /note= "KR1"
FT      misc_feature complement (39442..40609)
FT      /tag= aq
FT      /note= "DH1"
FT      misc_feature complement (40609..41842)
FT      /tag= ar
FT      /note= "AT1"
FT      misc_feature complement (41842..43093)
FT      /tag= as
FT      /note= "KS of extender module 1 (KS1)"
FT      misc_feature complement (43144..43660)
FT      /tag= at
FT      /note= "ACP of loading domain"
FT      misc_feature complement (43777..44629)
FT      /tag= au
FT      /note= "EK of loading domain"
FT      misc_feature complement (44974..46573)
FT      /tag= av
FT      /note= "COA ligase of loading domain"
FT      CDS      46754..47784
FT      /tag= aw
FT      /note= "fkbA gene"
FT      CDS      47785..52272
FT      /tag= ax
FT      /note= "fklp gene"
FT      CDS      52275..71465
FT      /tag= ay
FT      /note= "fkbA gene"
FT      misc_feature 52362..53576
FT      /tag= az
FT      /note= "KS7"
FT      misc_feature 53577..54716
FT      /tag= ba
FT      /note= "AT7"
FT      misc_feature 54717..55871
FT      /tag= bb
FT      /note= "KR2"
FT      misc_feature 56019..56819
FT      /tag= bc
FT      /note= "ER7"
FT      misc_feature 56943..57575
FT      /tag= bd
FT      /note= "KR7"

```


PF 13-MAY-1999; 99WO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085537.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J,
 PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A,
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I,
 PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
 DR WPI: 2000-126369/11.
 XX
 PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PS
 PS Claim 1: Page 994; 1097pp: English.
 XX
 XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 XX Sequence 1000 BP; 21 A; 53 C; 574 G; 22 T; 330 other;
 SO
 Query Match 10.4%; Score 47.2; DB 21; Length 1000;
 Best Local Similarity 34.3%; Pred. No. 0.097;
 Matches 142; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

DB 382 CCCCCCNCNCNCCCCCCCCCCCCCNCNNCNCNNCNCNCGCCGCCNC 329
 RESULT 14
 AA217038/C
 ID AA217038 standard; cDNA: 1454 BP.
 XX
 XX AA217038;
 DT 12-OCT-1999 (first entry)
 XX
 XX Human gene expression product cDNA sequence SEQ ID NO:4508.
 DE Human: gene: gene expression product; diagnosis: therapy: probe;
 KW detection: mapping: tissue typing: profiling: forensic: cancer;
 KW genetic analysis: colorectal cancer; breast cancer; lung cancer; ss.
 XX
 XX Homo sapiens.
 OS
 OS W09938972-A2.
 PN
 PN 05-AUG-1999.
 PD
 PD 28-JAN-1999; 99WO-US01619.
 PF
 PF 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
 PI Jones LM, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 XX WPI: 1999-494092/11.
 DR
 XX
 XX Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PS
 PS Claim 1: Page 2137; 2479pp: English.
 XX
 XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purposes, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 CC
 XX
 XX Sequence 1454 BP; 87 A; 112 C; 774 G; 183 T; 298 other;
 SO

```
Query Match      10.4%; Score 47.2; DB 20; Length 1454;
Best Local Similarity 40.4%; Pred. No. 0.099;
Matches 153; Conservative 0; Mismatches 225; Indels 1; Gaps 1;

QY 71 caccgcgagcgccggaagcgacgtcgccggaagcgctcaatctgctccgacgagc 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1420 CCNCNCGNCGCGCCACACCCACACACACCCCNACCCCTCCNCGCGCCCGCCNCGTC 1361
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 tggccgctgacgcccagcggtgatgagctggtctacaccccgagcgagcctcaccgaca 190
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1360 GNCACCTCCNNNNNCGCGGCACACGNCGCCATCCCGCCCGCCGACACACGACGCGCN 1301
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 cgaagcgcccgcaaaacgagcgtacacccctgagcaaacagcaatagcagcgatgcaag 250
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1300 CCCCNCNCGCGCCGACANNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 1241
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 gctaagtggtctacgtgaccccccaagtcgcgggcgacaccttgaagcgctgtagcacaac 310
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1240 CACTTNGCCNCTCGNCCGNCNNCCCGCCACGCACANAGCATCNCACCTCC 1181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 tggcgcccccgcgagcacaaccccccgagcaccacaccccggtcatcgacacaccccc 369
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1180 GGTGTCGCCCGCACGACGACGCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 1121
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 gatcgcgccgcatcgacgcgacaccccgagcagcagcagcagcagcagcagcagcagc 429
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1120 ACACCCACACCCCGCCACNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 1061

QY 430 ctggcggggtgctgcgcgc 448
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1060 CCNCACNCGCGCGCACNC 1042
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AAZ17263
ID AAZ17263 standard; cDNA; 1337 BP.
XX
AC AAZ17263;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN W0938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
XX
PR 28-JAN-1998; 98US-0072910.
XX
PR 24-FEB-1998; 98US-0075954.
XX
PR 31-MAR-1998; 98US-0080114.
XX
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI: 1999-494092/41.
XX
PT Novel human genes and their expression products which are
```

```
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2250-2251; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 1337 BP; 42 A; 577 G; 27 C; 22 T; 669 other;

Query Match      10.3%; Score 46.6; DB 20; Length 1337;
Best Local Similarity 29.5%; Pred. No. 0.13;
Matches 124; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 34 ccttlttgcacactgcgcgcgcggtggatgtgtccacccgacgagcgccgacgagcg 93
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 322 cnnncnncnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncc 381
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 94 acctgcggycaaaugcctcaatctgctccgacgagctggccgctacgcccagcggtc 153
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 cnnnnccnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 441
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 154 atggaactggtatcacacccgagcggtgagctcacgacacgacgagcgccgcaaacggca 213
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 442 cncncncnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncc 501
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 214 tcacctgagcaacacgacgaatagacagcagatgtcacgggctaagtggctacctgaccccc 273
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 502 ncncccccccccccccccccccccccccccccccccccccccccccccccccccccc 561
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 aagtgcgagcagcgtctggaagcgtgctgagcgaactggcgcccccgagcgagcgaac 333
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 562 cccccnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncc 621
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 334 cccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 393
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 622 cncncncnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncc 681
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 394 acccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 682 nncnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncc 741
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: April 2, 2002, 22:49:09
Job time: 9209 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:46:33 ; Search time 165.4 Seconds
(without alignments)
620.281 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcggagcgacacatca.....ccggctgcgcgctgac 453

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	408.4	90.2	1535	US-08-464-052-1	Sequence 1, Appli
2	408.4	90.2	1535	US-08-461-002-1	Sequence 1, Appli
3	408.4	90.2	1535	US-08-689-411-1	Sequence 1, Appli
4	408.4	90.2	1535	PCT-US94-09863-1	Sequence 1, Appli
5	409.4	90.2	4411529	US-09-103-840A-1	Sequence 1, Appli
6	366.8	81.0	12412	US-08-390-878-18	Sequence 18, Appli
7	351	77.5	650	US-08-464-052-5	Sequence 5, Appli
8	351	77.5	650	US-08-461-002-5	Sequence 5, Appli
9	351	77.5	650	US-08-689-411-5	Sequence 5, Appli
10	52.4	11.6	152331	US-09-128-155-16	Sequence 16, Appli
11	49.4	10.9	4403765	US-09-103-840A-2	Sequence 2, Appli
12	48.6	10.7	4403765	US-09-103-840A-2	Sequence 2, Appli
13	44.6	9.8	6085	US-09-029-603-4	Sequence 4, Appli
14	43.6	9.6	4411529	US-09-103-840A-1	Sequence 1, Appli
15	43.4	9.6	1028	US-08-118-200-1	Sequence 1, Appli
16	43.4	9.6	1028	US-08-458-745-1	Sequence 1, Appli
17	43.2	9.5	2712	US-09-025-691-4	Sequence 4, Appli
18	42.6	9.4	1018	US-08-444-083-6	Sequence 6, Appli
19	42.6	9.4	1018	US-08-286-304-6	Sequence 6, Appli
20	42.6	9.4	1018	US-08-442-745-6	Sequence 6, Appli
21	42.6	9.4	1018	US-08-443-129-6	Sequence 6, Appli
22	42.6	9.4	1018	US-08-443-952-6	Sequence 6, Appli
23	42.6	9.4	1018	US-08-443-130-6	Sequence 6, Appli
24	42.6	9.4	1018	US-08-898-911-6	Sequence 6, Appli
25	42.6	9.4	1018	PCT-US93-04467-6	Sequence 6, Appli
26	42.6	9.4	1203	US-09-086-010-1	Sequence 1, Appli
27	41.8	9.2	1140	US-09-105-537-15	Sequence 15, Appli

28	41.8	9.2	3292	3	US-09-320-878-22	Sequence 22, Appli
c 29	41.8	9.2	3765	3	US-07-705-490-1	Sequence 1, Appli
c 30	41.8	9.2	3765	4	US-07-751-891B-1	Sequence 1, Appli
c 31	41.8	9.2	4362	2	US-08-455-073A-1	Sequence 1, Appli
c 32	41.8	9.2	13613	4	US-09-105-537-3	Sequence 3, Appli
c 33	40.6	9.0	882	4	US-08-818-112-138	Sequence 138, App
c 34	40.4	8.9	530	3	US-08-758-662-4	Sequence 4, Appli
c 35	40.4	8.9	897	4	US-09-434-288-6	Sequence 6, Appli
c 36	40.4	8.9	4524	2	US-08-845-998-7	Sequence 7, Appli
c 37	40.4	8.9	4524	3	US-09-206-537-7	Sequence 7, Appli
c 38	40.4	8.9	4524	4	US-09-430-854-7	Sequence 7, Appli
c 39	39.6	8.7	666	2	US-08-875-034A-1	Sequence 1, Appli
c 40	39.6	8.7	1294	4	US-09-312-038-4	Sequence 4, Appli
c 41	39.6	8.7	2289	4	US-09-312-038-3	Sequence 3, Appli
c 42	39.6	8.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
c 43	39.4	8.7	2064	1	US-08-341-428-1	Sequence 1, Appli
c 44	39.4	8.7	11219	1	US-07-642-734C-1	Sequence 1, Appli
c 45	39.4	8.7	11219	3	US-08-439-009A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-464-052-1
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 40,127
; REFERENCE/DATA KEY NUMBER: 19605/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-1

Query Match 90.2%; Score 408.4; DB 3; Length 1535;
Best Local Similarity 97.8%; Pred. No. 3.3e-80;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;
Qy 1 gatcggagcgacacatcaaaagtatttcgcccacacatgcgcgcg 60
|||||
Db 827 GATCGCGGAGCGCACGCTAAAGTATTCGCGCCCTTTTCGCCACCTGCGCGCGG 886

Qy	61	tgaatgtgtcaaccgccaggcgccgaagcgacctgccggcaaaagcc--ctcaatatc	118
Db	887	TGGATGTGTCCACCCGCCAGCGCGCGAAGCGAGCTTGGCGGCAAAAGCGCTCAATATC	946
Qy	119	gtccgaagaatgtgcccgtactacgccagcgggtcatgagctggtacaccccaagcg	178
Db	947	GTCCCGACGAGCTGGCCCGCTACGCCAGCGGTCATGGACTGCTACACCCGACCGCG	1006
Qy	179	acctcacgcagacgaagcgcgcgcaaaagc'-gcatacccttgagcaaccagcaatag	237
Db	1007	ACCTCACGCACACGAACGCGCGCAACGCGGATCACCTTGAGCAACAGAAATACG	1066
Qy	238	acggcatgtcaagctgaatgtgctacctgaccccccaagtgcgaggcgcaacttgaagcc	297
Db	1067	ACGGCATGTCACGGCTAAGTGGCTACCTAGCCGCCCAAG-GCGGGGCACCTTTGAAGCC	1125
Qy	298	gtg'tagcaaaactgcccgcgcgcgcgaccaaaccgcgacgacacaccccaggtcatc	357
Db	1126	GTGCTAGCCAAATGGCCGCCGCCGGCGGCGAGCAACCCCGGACGACACCCCGGTGATC	1185
Qy	358	gacacacccccgatggccgcctatgaccgcgacaccccgagccaaagcccaacgaac	417
Db	1186	GACACACCCCGATGGCGCGCGATCGACCGGACACCCGACGCCAAGCCACGCAAC	1245
Qy	418	caagaaggctgtggccgggtgcgcgctgatac	453
Db	1246	CAGACGGGTGCTGGCCGGGCTGGCGCGCTGATC	1281

```

RESULT      2
US-08-461-002-1
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051

```

```

STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,002
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

```

Query Match 90.2%; Score 408.4; DB 4; Length 1535;
Best Local Similarity 97.8%; pred. No. 3.3e-80;

[illegible]

```

RESULT      3
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pete
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,411
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid

```

STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-689-411-1

Query Match 90.2%; Score 408.4; DB 4; Length 1535;
Best Local Similarity 97.8%; Pred. No. 3.3e-80;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

```
QY 1 gatcgagcgacacataaagtgatcgccctttttccacacgcccgcggg 60
   |||||||
Db 827 GATCGGCGAGCGACGCTCAAGTGAATTCGCGCCCTTTTCGCCACCTGCCGCGCGG 886
   |||||||
QY 61 tggatgttcaccccgccgagcgccgagcactgcggcaaaaggc--ctcaatatic 118
   |||||||
Db 887 TGGATGTGTCCACCGCGAGCGCGGAAGCGACTGCGCGCAAGCGGCTCAATATC 946
   |||||||
QY 119 gtccgacgagctggcccgctacgcccagcggtatgagactggctacacccgacgag 178
   |||||||
Db 947 GTCCCGAGAGCTGGCCCGCTACGCCAGCGGGTCAATGGACTGGCTACACCCCGAGCGG 1006
   |||||||
QY 179 acctcaccgacacgaacgcgcccgcgaacgc--gcataccctgagcaaccagcaatag 237
   |||||||
Db 1007 ACCTCACCAGACCGAACCGCGCGCAAGCGGATACCTGTGACRACCGACCAATATC 1066
   |||||||
QY 238 acggcatgtaoagtgtaagtgtactacctgaccccccaagtcggggccacacctttgaagcc 297
   |||||||
Db 1067 ACGGCATGTCAAGTGAATGTGCTACCTGACCCGCCAAG--CGCGGCGCACCTTTCAAGCC 1125
   |||||||
QY 298 gtctagcaaaactgcgcccccgagcgacacaccccgagacacaccccggtatc 357
   |||||||
Db 1126 GTCTAGCCAAACTGCGCCCGCGCGGACCAACCCCGAGCACACCCCGGCTCATC 1185
   |||||||
QY 358 gacacacccccgatcgccgcccattcgacgcgacaccccgacgacgcccgaagcccaacgcaac 417
   |||||||
Db 1186 GACACACCCCGATGCGCGCGCCATCGACCGGACACCCGAGCGACCGCCAGCGCAAC 1245
   |||||||
QY 418 cagcagggctgctggccgggtgctgcgcgctgac 453
   |||||||
Db 1246 CAGCAGGGCTGCTGCGCGGGTGCAGCGGCTGATC 1281
   |||||||
```

RESULT 4

PCT-US94-09863-1
Sequence 1, Application PCT/US94/09863
GENERAL INFORMATION:
APPLICANT: Riley, Lee W.
TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
TITLE OF INVENTION: UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Michael L. Goldman
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09863
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman Mr., Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716)263-1000

TELEFAX: (716)-263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1

Query Match 90.2%; Score 408.4; DB 5; Length 1535;
Best Local Similarity 97.8%; Pred. No. 3.3e-80;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

```
QY 1 gatcgagcgacacataaagtgatcgccctttttccacacgcccgcggg 60
   |||||||
Db 827 GATCGGCGAGCGACGCTCAAGTGAATTCGCGCCCTTTTCGCCACCTGCCGCGCGG 886
   |||||||
QY 61 tggatgttcaccccgccgagcgccgagcactgcggcaaaaggc--ctcaatatic 118
   |||||||
Db 887 TGGATGTGTCCACCGCGAGCGCGGAAGCGACTGCGCGCAAGCGGCTCAATATC 946
   |||||||
QY 119 gtccgacgagctggcccgctacgcccagcggtatgagactggctacacccgacgag 178
   |||||||
Db 947 GTCCCGAGAGCTGGCCCGCTACGCCAGCGGGTCAATGGACTGGCTACACCCCGAGCGG 1006
   |||||||
QY 179 acctcaccgacacgaacgcgcccgcgaacgc--gcataccctgagcaaccagcaatag 237
   |||||||
Db 1007 ACCTCACCAGACCGAACCGCGCGCAAGCGGATACCTGTGACRACCGACCAATATC 1066
   |||||||
QY 238 acggcatgtaoagtgtaagtgtactacctgaccccccaagtcggggccacacctttgaagcc 297
   |||||||
Db 1067 ACGGCATGTCAAGTGAATGTGCTACCTGACCCGCCAAG--CGCGGCGCACCTTTCAAGCC 1125
   |||||||
QY 298 gtctagcaaaactgcgcccccgagcgacacaccccgagacacaccccggtatc 357
   |||||||
Db 1126 GTCTAGCCAAACTGCGCCCGCGCGGACCAACCCCGAGCACACCCCGGCTCATC 1185
   |||||||
QY 358 gacacacccccgatcgccgcccattcgacgcgacaccccgacgacgcccgaagcccaacgcaac 417
   |||||||
Db 1186 GACACACCCCGATGCGCGCGCCATCGACCGGACACCCGAGCGACCGCCAGCGCAAC 1245
   |||||||
QY 418 cagcagggctgctggccgggtgctgcgcgctgac 453
   |||||||
Db 1246 CAGCAGGGCTGCTGCGCGGGTGCAGCGGCTGATC 1281
   |||||||
```

RESULT 5

US-09-103-840A-1/r
Sequence 1, Application US/09/103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen K.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match

90.2%; Score 408.4; DB 4; Length 4411529;

TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-461-052-5

Query Match 77.5%; Score 351; DB 3; Length 650;
Best Local Similarity 97.7%; Pred. No. 7.2e-68;
Matches 388; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

QY 60 gtggatgtgtccaccgagcgccgagcgagcgtgctatgagcgtgctacaccccgagcgc 117
|||||
DB 1 GTGGATGTGTCCACCAGCGCGCGGCGGAGCGGACCTGGCGGCGGCAAGCGCTCAATAT 60
QY 118 cgtcccgagcagctgcccgtacgcccagcgggtcatgagcgtgctacaccccgagcgc 177
|||||
DB 61 CGTCCCAGCAGCTGCGCGCTACGCGCCGAGCGGGTCTGACTGGCTACACCCCGACGCG 120
QY 178 gacctcaccgacacgacgagcgcgcgcgcaaacgc-gcatcacctgagcagcagcagc 236
|||||
DB 121 GACCTCACCAGACCGACGCGCGCGGCGGCAAGCGGCGATCACCTGAGCAACCGCAATAC 180
QY 237 gacggcatgtcacggttaagtgtacctgaccccccaagtcgaggcgacaccccgagcgc 296
|||||
DB 181 GACGGCATGTACGGGTAAAGTGGCTACCTGACCGCGGCGGCGGCGGCGGCGGCGGCGG 239
QY 297 cgtgctagcgaactggtggtggtggtggtggtggtggtggtggtggtggtggtggt 356
|||||
DB 240 CGTGTAGCGCAACG 299
QY 357 cgacacaccccgatggtggtggtggtggtggtggtggtggtggtggtggtggtggt 416
|||||
DB 300 CGACACACCCCGGATGCG 359
QY 417 ccacacggcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 453
|||||
DB 360 CCACGACGGGCTGCTGGCGGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396

RESULT 8

US-08-461-002-5
Sequence 5, Application US/08461002
Patent No. 6214543
GENERAL INFORMATION:
APPLICANT: Riley M.D., Lee W.
TITLE OF INVENTION: DNA Molecule Encoding for Cellular
TITLE OF INVENTION: Uptake of Mycobacterium tuberculosis and uses thereof
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,002
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727

REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-461-002-5

Query Match 77.5%; Score 351; DB 4; Length 650;
Best Local Similarity 97.7%; Pred. No. 7.2e-68;
Matches 388; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

QY 60 gtggatgtgtccaccgagcgccgagcgagcgtgctatgagcgtgctacaccccgagcgc 117
|||||
DB 1 GTGGATGTGTCCACCAGCGCGCGGCGGAGCGGACCTGGCGGCGGCAAGCGCTCAATAT 60
QY 118 cgtcccgagcagctgcccgtacgcccagcgggtcatgagcgtgctacaccccgagcgc 177
|||||
DB 61 CGTCCCAGCAGCTGCGCGCTACGCGCCGAGCGGGTCTGACTGGCTACACCCCGACGCG 120
QY 178 gacctcaccgacacgacgagcgcgcgcgcaaacgc-gcatcacctgagcagcagcagc 236
|||||
DB 121 GACCTCACCAGACCGACGCGCGCGGCGGCAAGCGGCGATCACCTGAGCAACCGCAATAC 180
QY 237 gacggcatgtcacggttaagtgtacctgaccccccaagtcgaggcgacaccccgagcgc 296
|||||
DB 181 GACGGCATGTACGGGTAAAGTGGCTACCTGACCGCGGCGGCGGCGGCGGCGGCGGCG 239
QY 297 cgtgctagcgaactggtggtggtggtggtggtggtggtggtggtggtggtggtggt 356
|||||
DB 240 CGTGTAGCGCAACG 299
QY 357 cgacacaccccgatggtggtggtggtggtggtggtggtggtggtggtggtggtggt 416
|||||
DB 300 CGACACACCCCGGATGCG 359
QY 417 ccacacggcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 453
|||||
DB 360 CCACGACGGGCTGCTGGCGGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396

RESULT 9

US-08-689-411-5
Sequence 5, Application US/08689411
Patent No. 6224801
GENERAL INFORMATION:
APPLICANT: Riley M.D., Lee W.
APPLICANT: Chong, Belle
TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,411
FILING DATE:

[illegible]

```

RESULT 10
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Fan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TYPE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match      11.6%; Score 52.4; DB 3: Length 152331;
Best local Similarity 45.6%; Prod. No. 0.003;
Matches 182; Conservative 0; Mismatches 217; Indels 0; Gaps 0

QY 29 cagcgcccttttttcccccattgtcccccgcgcgtggatgtgtccaccgccagccgcgcga 88
DB 21884 ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc 21943

QY 89 aagcaacttgcgcgcgaagggtcaatctgtcccgacgagtgtgcccgcgtactgcgcagc 148
DB 21944 ccccggttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22003

QY 149 gggtcagtgaactggctacaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 208
DB 22004 cccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22063

QY 209 ggcacatccacttgcgaaccaggaataatgcgcgcgcgcgcgcgcgcgcgcgcgcgc 268
DB 22064 ccgggtccacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22123

QY 269 cccccaatttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 328
DB 22124 cccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22183

QY 329 ccaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 388
DB 22184 cccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22243

QY 389 gcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 427
DB 22244 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22282

RESULT 11
US-09-103-840A-2/v
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen K.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24 366, 2000/7-00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

	Query Match	10.98;	Score 49.4;	DB 4;	Length 4403765;
	Best Local Similarity	98.0%;	Pred. No. 0.018;		
	Matches 447;	Conservative	0;	Mismatches	5; Indels 4; Gaps 3;
OY	i gatcgcgagdgccacatcaaaagtattcgcgcccttttctgcgcacctgccccgcgcgg	60			
DG	1780338 GATTCGGCGAGGCCACATCAAACTGATTGCGCCCTTTTTCGCCACCTGCCGCCGCCG	1780279			

Qy 61 tggatgtgtccaccgcgacagcgccgaagcgcgacctgcccggcaaaaggc--ctcaatatc 118
| | | | | | | | | | | | | | | | | | | | | |
Db 1780278 TGGATGTGTCACCCGCCAGCGCCGAAGCGCACCTGGCGGCAAAAGCGCCTCAATATC 1780219

Qy 119 gtcccgacgagctggcccgctactagccagcggggtcatgtgaactggtcacacccccgacggcg 178
| | | | | | | | | | | | | | | | | | | | | |
Db 1780218 GTCCCCAGCAGCTGGCCCGCTACGCCACGGGTCTATGGACTGTGCCTACACCCCAGCGGCG 1780159

Qy 179 acctcacgcacacgaagcgccgcccgaagcg-gcatacccttgacaaccaaatcag 237
| | | | | | | | | | | | | | | | | | | | | |
Db 1780158 ACCTCACGCACACCGAACCGCGCCCGCAAACCGCGCATCACCTTGAGCAACACGAATACG 1780099

Qy 238 acggnatgtcacgctgaagtgtgttacctgaccccccaagtctcgcgggccacatttgaagcc 297
| | | | | | | | | | | | | | | | | | | | | |
Db 1780098 ACGCATGTACCGGCTAATGTGGTACTACTGACCCTCCCAAG-CGGCGGCGCACCTTTGAAGCC 1780040

Qy 298 gtgttagcctaactggccgcccccggcgcgaccaaaccgcagaccacacccccggtcatc 357
| | | | | | | | | | | | | | | | | | | | | |
Db 1780039 GTGCTAGCCAAACTGGCCGCCGCCCGCGCGGCGACCAACCCCGACGACACACCCCGGTATC 1779980

Qy 358 gacaccacccccgatgtcgggccgcocatcgacggagacaccccgagccaaagcccaacgcaac 417
| | | | | | | | | | | | | | | | | | | | | |
Db 1779979 GACACCACCCCGATCGCGCGGCATCGACCGGGACACCCGACGACCGCAAGCCCAACGCAAC 1779920

Qy 418 cagcagcgggtgtgtgcccgggtgtgcgcgctgac 453
| | | | | | | | | | | | | | | | | | | | | |
Db 1779919 CAGCAGCGGCTGTGTGCCCGGCTTGC CGCGGCTGATC 1779884

```

RESULT      12
US-09-103-840A-2
; Sequence 2, Application us/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-200007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

	Query Match	10.7%	Score 48.6;	DB 4;	Length 4403765;
	Best Local Similarity	97.1%;	Pred. NO. 0.027;		
	Matches 443; Conservative	0;	Mismatches	9;	Indels 4; Gaps 3;
QY	1 gatcgcgagcgccacatcaaatgattcgcgcgcccttttcgcccaactgccgcgcggcg 60				
	Db 3869789 gatcgcgagcgccaagtcaaagtattcgcgccttttcgcccaactgccgcgcggcg 3869848				
QY	61 tgaatgtgtccaccgcgcgagcgccgaagcgcacctgccgcgcaaaaggc--ctcataatc 118				
	Db 3869849 tgagtggtccaccgcgcgagcgccgaagcgcacctgccgcgcaaaaggcctcaatc 3869908				
QY	119 gtcccagcagtggtgcgcgtacgcgccagcggtgatgtgactggctacccccgcgcggcg 178				
	Db 3869909 gtcccagcagtggtgcgcgtacgcgccagcggtcatgagcttggctacccccgcgcggcg 3869968				

[illegible]

```

RESULT 13
US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Rietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; CURRENT FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085

```

	Query Match	9.8%	Score 44.6;	DB 4;	Length 6085;
	Best Local Similarity	51.2%;	Pred. No. 0.11;		
	Matches 104;	Conservative 0;	Mismatches 99;	Indels 0;	Gaps 0;
Qy	249	cgggtaagtgcacctgaaccccccaagtcgcgggccacctttgaagccgctagaccaa	308		
Db	5298	cqacgaagtatcacqqtgtccaacaccgcgcccccacccatctatgcacatgcactccac	5357		

QY 309 actgagcgcccgcgagcaaccacgcagacacacccgggtcatcgacacacccc 368
 3538 cgagcgcaaccccgctctcgacgctcgagagacgtctcatgagacagagcca 5417
 QY 369 cgatcgcgccgcctatcgaccccgacacccgacgacgcaacgaacacacgacggtc 428
 5418 ggtcgagcgccgcctacgacacacccgtgctgctccggtccactgacggcca 5477
 QY 429 gctggccgggctgcgcgcgtga 451
 Db 5478 gtgcgtcgacatggcgccgcgtga 5500

```

RESULT 14
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match          9.68; Score 43.6; DB 4; Length 4411529;
Best Local Similarity 97.8%; Pred. No. 0.32;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

QY      1  gatcgcgagcgcacatcaagtgttcgcgccttttcgccacctgccgcgcgcg 60
Db      3883895  gatcgcgagcgacgtcaagtgttcgcgccttttcgccacctgccgcgcgcg 3883954

QY      61  tggatgtgtcacccgcagcgccgaaagccgacgttcggcgcaagggc -ctcaatatc 118
Db      3883955  tggatgtgtcacccgcagcgccgaaagccgacgttcggcgcaagggcctcaatatc 3884014

QY      119  gtcccgcagcagctgtggccgcgtacgccacgcgggtcctagagctggctaccctccgcgcgc 178
Db      3884015  gtcccgcagcagctgtggccgcgtacgccacgcgggtcctagagctggctaccctccgcgcgc 3884074

QY      179  aactcacgcacacgcagcgcgcgcgcaaacgc -gcatcaccttgagcaaccgcaatacg 237
Db      3884075  aactcacgcacacgcagcgcgcgcgcaaacgcgcgcctgagcaaccgcaatacg 3884134

QY      238  acggcatgtcagcgctaagtggctacctgaccccccaagtctgcggcgccacctttgaagcc 297
Db      3884135  acggcatgtcagcgctaagtggctacctgaccccccaagtctgcggcgccacctttgaagcc 3884193

QY      298  gtgctagcacaactggccgcccccgcgcgacacaccccgacgacacaccccggtcatc 357
Db      3884194  gtgctagcacaactggccgcccccgcgcgacacaccccgacgacacaccccggtcatc 3884253

QY      358  gacacacccccgatgcggcgccatctgacccgcgacacccgcagccaagcccaacgcaac 417
Db      3884254  gacacacccccgatgcggcgccatctgacccgcgacacccgcagccaagcccaacgcaac 3884313

QY      418  cacgacgggctgtgcgcgggctgcgcgcgtgac 453
Db      3884314  cacgacgggctgtgcgcgggctgcgcgcgtgac 3884349

```

```

RESULT 15
US-08-118-200-1/c
: Sequence 1, Application US/08118200
: Patent No. 6197500
: GENERAL INFORMATION:
: APPLICANT: SUTHERLAND, Grant R
: APPLICANT: RICHARDS, Robert I
: APPLICANT: SCHLESSINGER, David
: APPLICANT: NAGARAJA, Kamalah
: APPLICANT: KREMER, Eric J
: APPLICANT: YU, Shi
: APPLICANT: BAKER, Elizabeth
: APPLICANT: MULLEY, John C
: APPLICANT: MANDEL, Jean-Louis
: APPLICANT: PRITCHARD, Melanie April
: APPLICANT: LYNCH, Michael
: TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
: TITLE OF INVENTION: FRAGILE X SYNDROME
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWICKER & MATIUS
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22314-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentLin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/118,200
: FILING DATE: 09-SEP-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/802,650
: FILING DATE: 05-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,232
: FILING DATE: 20-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/638,518
: FILING DATE: 04-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/966,517
: FILING DATE: 23-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Crane, Fentley, Sharon E
: REGISTRATION NUMBER: 46,114
: REFERENCE/DOCKET NUMBER: 020160-164
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1028 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-118-200-1

```

	Query Match	9.6%	Score 43.4;	DB 4;	Length 1028;
	Best Local Similarity	53.9%;	Pred. No. 0.17;	Mismatches	89; Conservative
		0;	Mismatches	76; Indels	0; Gaps
QY	280	cgggcccacatttgaagccgtgcctagccaaactgcgcccccccgccgacccaaccgccac	339		
DG	478	CGGCGCCGCCCGCCGCAGCGCAGCGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC	419		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 21:47:49 ; Search time 3835.7 Seconds
(without alignments)
1269.087 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccggcgtcgcgcgctgac 453

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estro:*
8:	em_estov:*
9:	em_htc:*
10:	gb_estl:*
11:	gb_est2:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pin:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	61.2	13.5	1014	13	CNS0057B	AL057121 Drosophil
2	59.6	13.2	1101	13	CNS014PA	AL104440 Drosophil
3	58.2	12.8	576	11	BG786264	BG786264 SEAMC006
4	57.4	12.7	1101	13	CNS0153F	AL104949 Drosophil
5	56.8	12.5	520	11	BG241147	BG241147 OVI_38_G1
6	56.6	12.5	848	13	CNS021YE	AL220271 Tetraodon
7	56.4	12.5	925	13	CNS0091P	AL053013 Drosophil
8	56.4	12.5	1101	13	CNS017SY	AL108460 Drosophil
9	56.2	12.4	644	13	CNS06094	AL1393518 T7 end of
10	56	12.4	632	11	BG320266	BG320266 Zm03_0390
11	56	12.4	812	10	AL537183	AL537183
12	56	12.4	921	13	AZ211117	AZ211117 SP_0159_B

C 13	55.8	12.3	892	13	AZ200928	AZ200928 SP_0134_B
14	55.6	12.3	786	13	AZ196321	AZ196321 SP_1031_B
C 15	55.4	12.2	566	13	CNS03JN0	AL247077 Tetraodon
16	55.4	12.2	776	13	CNS010RY	AL099352 Drosophil
17	55.4	12.2	848	13	CNS0118A	AL099940 Drosophil
18	55.4	12.2	908	13	CNS006B4	AL064031 Drosophil
19	55	12.1	839	13	CNS004NB	AL054280 Drosophil
20	55	12.1	1016	11	BG368625	BG368625 HVSME1002
C 21	54.8	12.1	878	11	BG786339	BG786339 SEAMC006
22	54.8	12.1	1038	13	CNS015VV	AL105973 Drosophil
23	54.4	12.0	631	11	BG413362	BG413362 7o37d09_X
C 24	54.4	12.0	666	11	BG786337	BG786337 SFAUMC006
C 25	54.2	12.0	745	13	AZ046300	AZ046300 ubeu0091L
C 26	53.8	11.9	892	13	AZ194728	AZ194728 SP_1028_B
C 27	53.6	11.8	699	11	BG076530	BG076530 H3001D05-
C 28	53.6	11.8	895	13	CNS0071A	B09456 F26A17-Sp6
C 29	53.6	11.8	978	13	B09456	AL066286 Drosophil
C 30	53.6	11.8	999	10	AL525224	AL525224
C 31	53.2	11.7	549	13	CNS04CNH	AL284678 Tetraodon
C 32	53	11.7	935	13	CNS006XK	AL066051 Drosophil
33	52.8	11.7	794	13	AZ185894	AZ185894 SP_1005_B
34	52.6	11.6	1028	13	CNS015AB	AL105197 Drosophil
C 35	52.4	11.6	673	13	AZ184272	AZ184272 SP_1002_B
C 36	52.4	11.6	741	13	AZ194927	AZ194927 SP_1029_A
37	52.4	11.6	775	13	AZ185516	AZ185516 SP_1005_A
38	52.4	11.6	795	13	AZ188474	AZ188474 SP_1012_A
C 39	52.4	11.6	802	13	CNS02R43	AL210108 Tetraodon
C 40	52.4	11.6	1016	13	CNS03JN9	AL249930 Tetraodon
C 41	52.4	11.6	1101	13	CNS0181P	AL108775 Drosophil
42	52.2	11.5	915	13	AZ195437	AZ195437 SP_1030_A
43	52.2	11.5	925	10	AL581447	AL581447 Drosophil
C 44	52.2	11.5	932	13	CNS0072Q	AL066742 Drosophil
C 45	52.2	11.5	1059	13	CNS0155J	AL105025 Drosophil

ALIGNMENTS

RESULT 1
CNS0057B
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS0057B 1014 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACRILKil of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL057121
AL057121.1 GI:4932509
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source 1..812
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODF019YL14"
 /clone_lib="LVI_FLO13_FBRn1"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /notes=Organ: Fetal brain; Vector: pCMVSPORT 6; lnt strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 82 a 102 c 14 g 69 t 45 others

ORIGIN

Query Match 12.4%; Score 56; DB 10; Length 812;
Best Local Similarity 43.3%; Pred. No. 0.14;
Matches 182; Conservative 14; Mismatches 224; Indels 0; Gaps 0;

QY 29 cgcgcccttttgcgccacctgcccgcggcgttgatgtgtccacccgagcaggcgccga 88
 | ||||| | |||-||| | ||||| | ||||| | ||||| | |||||
Db 511 CCCCCCCCCTCCCGCCCCCCTCCCGCCCCGCCCCGCCGGCCGCCTCCCGCCCCCA 452

QY 89 agcggacctgcgggaaaggcctaataatgttcaccacgagtgcgcgatgcactaac 148
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 451 CGCGCGCCCCCCCCCCCCCCG 392

QY 149 gggtcatgactggttacacccgcgacgcgcgcaacctcacccacacgcgccccgacaac 208
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 391 CCCCCCMAGCMCCCCCCCCCCCCCCCCCCCCACCCCGCCCCCGCCCCCCCCCCCC 332

QY 209 gcgcataccctgagcaacacagaataatcacgacggcatgtcacggtaagtggctaccgtac 268
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 331 MCACAGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCCC 272

QY 269 cccccaagtgcggggccacettttaagcgcgtgtagcgaacctggcgcgcccccgacgcta 328
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 271 CCCCCCGCCM 212

QY 329 caaaccccgacacacacacaccccggttcacgacacacaccccgatggcgcygcraaaagc 308
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Ddb 211 CCCACGMSMCMCCCCSCG 152

QY 389 ggcgacccgcgagccaagcccacacgcaacacgcgacggcgtgctgcgaggggtgcgcgcq 448
 | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Ddb 151 CTCTCCCGCCGMACTCCCGCCCGCACGCGCCCGSGCCSCCCSCCCCGMMAMGCCCCGCGCG 92

RESULT 12 A2211117
LOCUS SP_0159_B1_C05_SP6E.Strongylocentrotus purpuratus, purple sea 31-AUG-2000
DEFINITION Strongylocentrotus purpuratus, genomic clone Plate-159 Col-9 Row-F, DNA sequence.
A2211117
KEYWORDS Strongylocentrotus purpuratus.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euarchinozoa; Echinacea; Echinoidea;

<hr/>						
ACCESSION	AZ200928					
VERSION	AZ200928.1	GI:8395946				
KEYWORDS	GSS.					
SOURCE	Strongylocentrotus purpuratus.					
ORGANISM	Strongylocentrotus purpuratus					
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;					
	Echinoidea; Euclinoidea; Echinacea; Echinoidea;					
	Strongylocentrotidae; Strongylocentrotus.					
REFERENCE	1 (bases 1 to 892)					
AUTHORS	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.					
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources					
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)					
MEDLINE	20402566					
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 134 row: F column: 24 Seq primer: T7 Class: BAC ends High quality sequence stop: 892. Location/Qualifiers . .892 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="Plate=134 Col=24 Row=F" /clone.lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library." /note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli DH10B"					
FEATURES	source					
	BASE COUNT	47 a	175 c	542 g	119 t	9 others
	ORIGIN					
	Query Match	12.3%;	Score 55.8;	DB 13;	Length 892;	
	Best Local Similarity	46.08;	Pred. No. 0.15;	Mismatches 0;	Gaps 0;	
	Matches 189;	Conservative	0;	Indels 222;		
QY	29	cgcgcccttttggccacactgcccgcgggttgatgtgtccaccgcagcgccgcga	88			
Db	778	CCCGCGGCCACCCACACCCCCCGCGCCCGCCGCCCGCCCGCCCGCCAGCCGCAGGGCGCCGCC	719			
QY	89	agcgcacctgcggcaaggcctaataatcgtccgcagcagagtggcccttatccttcac	148			
Db	718	ACCACCGCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCA	659			
QY	149	gggtcatggaactggtcacaccgcagcggcagctcacgcacacccaagcgcgcgaaac	208			
Db	658	CCGCGCCCGCCCGCCCGCAGACGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC	599			
QY	209	gcgcataccctgagcaacagacaatacagcgcgatgcagcgttaagtgccttac	268			
Db	598	CCGACCCGCGCCCGCGCCCGCACCCCGCAGCCACCGCGCGCCCGCCCGCCCGCCCGCCAGCCG	539			
QY	269	ccccaaagtgcggggccacattttaagccgtgctagccaaactggcgcccccggcgcgga	328			
Db	538	CCCACCGCCCGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG	479			
QY	329	ccaaacccgcagcaccacaccccggtcatcgacaccaccocccgatgcggcgccatcgacc	388			
Db	478	CCACCGCCCGCCCGCAGCCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	419			
QY	389	gcgacacccgcagcccaagcccaacacacacacgacgggtgctggcgccggcg	439			
Db	418	CCGCGCGCGCCCGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC	368			

RESULT	14
AZ196321	
LOCUS	786 bp DNA 31-AUG-2000
DEFINITION	SP_1031_B2_D02_17A Strongylocentrotus purpuratus, purple sea urchin clone Plate-1031 Col-4 Row-H, DNA sequence.
ACCESSION	AZ196321
VERSION	AZ196321.1 GI:8390144
KEYWORDS	GSS.
SOURCE	Strongylocentrotus purpuratus.
ORGANISM	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euchinoidea; Echinacea; Echinoidae; Strongylocentrotidae; Strongylocentrotus l (bases 1 to 786)
REFERENCE	Cameron,R.A., Mahairas,G.C., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
AUTHORS	A sea urchin genome project: Sequence scan, virtual map, and additional resources
TITLE	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000) 20402566
JOURNAL	Contact: Cameron, RA, Davidson, EH, Hood, L
MEDLINE	Division of Biology 156-29
COMMENT	California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: dcameron@caltech.edu Plate: 1031 row: H column: 4 Seq primer: 'T' Class: BAC ends High quality sequence stop: 786. Location/Qualifiers 1..786 organism="Strongylocentrotus purpuratus" db_xref="taxon:7668" clone="plate-1031 col=4 Row-H" clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" note="Organ: sperm; Vector: BAC#3.6; BAC Clones in E-Coli DH10B"
FEATURES	source
BASE COUNT	49 a 413 c 171 g 80 t 73 others
ORIGIN	
Query Match	12.2% Score 55.6; DB 13; Length 786;
Best local Similarity	47.2%; Prod. No. 0.1b;
Matches 189:	Conservative 0; Mismatches 210; Indels 1; Gaps 1;
Qy	50 qccgcgagcgatgatggcttcacgtcagcgccgcgaagcgacctgcggcaagc 109
Dd	285 GCCCGCCGGGGGGGGGGNGCCTTCCTCCCCGCCGCCGCCGCCGCCGCCGCC 344
Qy	110 ctcaatatctcgcagcgactggcccgcctacgccagcggttcatggaatggctacac 169
Dd	345 CCCCCTCCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 404
Qy	170 ccgacggcgacctcacgaccgaacgagcgccgcgaagcgcatcaccttgagcaaca 229
Dd	405 CCCCCNCCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 464
Qy	230 gcaatacagc-ggcattgtcacggctaaagtggctacctgaccccccaagtcggggccaac 288
Dd	465 CCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCGC 524
Qy	289 tttagaacgctgctagccaacttgcgcgcccgccgcgacaccaacccccagacacac 348
Dd	525 CCCCCCGCCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCGC 584
Qy	349 ccggttatctatgacacaccccccgatgcggcgcccatcagccqgcagaccccgacac 408

